ESTATÍSTICA DESCRITIVA

Nilo Antonio de Souza Sampaio
Alzira Ramalho Pinheiro de Assumpção
Bernardo Bastos da Fonseca

Editora Poisson

Nilo Antônio de Souza Sampaio
Alzira Ramalho Pinheiro de Assumpção
Bernardo Bastos da Fonseca

Estatística Descritiva

1ª Edição

Belo Horizonte
Poisson
2018

Nilo Antônio de Souza Sampaio

- Doutor em Engenharia Mecânica pela UNESP-FEG - 2011.
- Professor Adjunto de Probabilidade e Estatística UERJ-FAT, Professor Associação Educacional Dom Bosco (AEDB) e Professor Faculdade Sul Fluminense (FASF).
- Trabalha com Matemática Aplicada, Probabilidade e Estatística, trabalhando nas seguintes áreas: Pesquisa sobre Aplicações da Estatística e da Matemática em todas as Ciências e Planejamento de Experimentos

Alzira Ramalho Pinheiro de Assumpção

- Doutora em Engenharia de Produção pelo Instituto Luiz Alberto Coimbra de Pós Graduação e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro – PEP/ COPPE/UFRJ – 1996.
- Professora Associada UERJ-FAT.
- Trabalha com Pesquisa Operacional, Raciocínio Lógico, Negociação, Metodologia do Ensino Superior.

Bernardo Bastos da Fonseca

- Doutor em Engenharia de Produção pelo Instituto Luiz Alberto Coimbra de Pós-Graduação e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro - PEP/COPPE/UFRJ - 2014.
- Professor Adjunto UERJ-FAT.
- Trabalha com pesquisa na área de segurança, meio ambiente e saúde.

PREFÁCIO

Estamos colocando a disposição dos colegas professores e aos interessados em estatística de modo geral este livro de Estatística Descritiva. O conteúdo deste livro apresenta os conceitos básicos iiniciais de um curso de estatística. Atendendo a uma crítica constante de professores e alunos com respeito aos conceitos de estatística que são abordados sempre do ponto de vista matemático, com ênfase em cálculos e medidas procuramos dar um enfoque maior à parte conceitual procurando realizar a formação com base no conhecimento e as interpretações do fenômeno estatístico.

DEDICATÓRIA

Sendo o Autor principal deste material, gostaria de dedicar este livro a meu pai Emanuel Pedro W Sampaio (in memorian) que foi meu primeiro mestre e incentivador na matemática e nos estudos em geral e por quem eu guardo uma grande saudade e admiração e ao meu grande amigo Cel.Prof. Ângelo Castro (in memorian) que foi meu grande mestre e exemplo de professor.

Nilo Sampaio

SUMÁRIO

Conceitos Importantes de Estatística

1.1 CONCEITOS INICIAIS

A Estatística é a ciência matemática que trabalha com dados comparativos e tem sido evidenciada atualmente pelas pesquisas realizadas por institutos com as mais diversas finalidades, tais como, consumo, eleitorais, renda, etc. Os dados são levantados, organizados e analisados e, com base nos mesmos, são feitasas projeções desejadas. Para isso, existem critérios que são obedecidos e métodos que são utilizados para a obtenção dos dados e sua organização em tabelas e gráficos quepossibilitem sua análise.

Com o avanço das novas tecnologias, temos uma diversificação importante e criativa por partes dos profissionais ligados à informação. Para acompanhar essa evolução e para que existam condições de analisar e criticar as informações obtidas são utilizados os métodos estatísticos.

De modo geral, observamos que a palavra tem dois significados básicos. No primeiro sentido, o termo é usado em relação a números específicos obtidos de dados, conforme ilustrado nos exemplos seguintes:

- Em pesquisa feita pelo instituto Vox Populi junto à 500 pessoas, foi formulada uma pergunta sobre a intenção do Movimento dos Trabalhadores Sem Terra - MST. 20 % responderam que o MST quer apenas fazer agitação política. (Revista Veja, Setembro de 2003);
- Um grupo de cientistas divulgou a taxa de mortalidade para algumas epidemias. A taxa de mortalidade da varíola é de 30 % (Revista Veja, Maio de 2003);
- Foram eleitos para a Câmara de Deputados em Brasília em outubro de 2002 o total de 70 deputados do PMDB.
- O outro sentido se refere à estatística como método de análise.
- Em 1662, John Graunt publicou informes estatísticos sobre nascimentos e mortes. O trabalho de Graunt foi secundado por estudos de mortalidade, tamanho de populações, rendas e taxas de desemprego;
- As taxas de desemprego, de inflação e os índices do consumidor são calculados a intervalos regulares e servem de base para que empresários tomem decisões que afetam a futura contratação de empregados, níveis de produção e expansão para novos mercados.

1.2 ORIGEM E CONCEITO DA ESTATÍSTICA

Estatística é a parte da Matemática que utiliza métodos científicos para coleta, organização, apresentação e análise de dados, bem como a obtenção de conclusões e tomada de decisões baseadas na análise dos dados.

É uma coleção de métodos para planejar experimentos, obter dados e organizá-los, resumi-los, analisá-los, interpretá-los e deles extrair conclusões.

1.3 DIVISÃO DA ESTATÍSTICA

- **Estatística Descritiva:** é aquela que tem por finalidade o fenômeno estatístico onde há coleta, organização e apresentaçãodos dados obtidos.

- **Estatística Indutiva:** é também chamada Estatística Amostral. É aquela que a partir de algumas informações e mediante métodos adequados estima, induz e conclui possíveis resultados sobre o fenômeno estudado.

- **Probabilidade:** é o ramo da estatística utilizado na análise de situações em que envolvam o acaso.
  - Ex: jogo de dados, lançamento de moedas.

- **Inferência Estatística ou Estatística Indutiva:** é a área que diz respeito á análise e interpretação de dados amostrais – constitui o conjunto de métodos para tomar decisões em situações onde há incerteza ou variação.
  - Ex: suponha que o diretor ou gerente de uma grande empresa desejasse levantar o grau de satisfação dos seus empregados. A população seria todos os trabalhadores da empresa. Seleciona-se uma parte desses empregados para responder à uma pesquisa, constituindo-se uma amostra. O objetivo do trabalho é levantar um determinado parâmetro da população. Isso seria alcançado utilizando-se dados obtidos da amostra de trabalhadores (estatística) para estimar atitudes ou características de interesse da população estudada.

As áreas da Estatística não são separadas ou distintas. Ao contrário, elas tendem a se entrelaçar. Descrever e resumir dados constitui a primeira fase de sua análise. A teoria da amostragem e da inferência estatística baseiam-se na teoria da probabilidade.

1.4 DEFINIÇÕES:

- **População:** coleção completa de todos os elementos (valores, pessoas, medidas etc) a serem estudados.

  Ex. Todos os alunos da universidade, A cidade de Bom Sorriso possui 17 bairros, torcedores do Fluminense.

Divide-se em:

a) População Finita: quando o número de elementos pode ser conhecido.

Ex: O Congresso Nacional brasileiro possui 513 deputados federais.

b) População Infinita: quando não se conhece o número de elementos.

Ex: O público presente no comício foi estimado pela polícia militar em 20 mil pessoas.

➡ Amostra: é qualquer subconjunto não vazio da população, é uma subcoleção de elementos extraídos de uma população.

Ex.: Para a realização de uma pesquisa de satisfação foram distribuídos questionários a 10% dos alunos de cada curso da universidade.

➡ Censo: é o estudo estatístico realizado em toda a população.

Ex.: O censo realizado em 2010indica que a população brasileira é de 190.755.799 habitantes (IBGE)-2010.

➡ Parâmetro: é uma medida numérica que descreve uma característica de uma *população.*

Ex: A altura média do brasileiro é de 1,68 metros.

➡ Estatística: é uma medida numérica que descreve uma característica de uma *amostra.*

Ex: Em uma amostra de 200 brasileiros verificou-se que a altura média é de 1,70 metros.

➡ Variável: é o conjunto de resultados possíveis de um fenômeno. Nas várias maneiras de classificarmos os conjuntos de dados, veremos diferentes tipos de variáveis, como por exemplo:

- altura em centímetros
- peso em Kg
- distância em metros
- volume em m^3
- cor dos olhosem preto, castanho, azul, etc

Dado: É um valor *numérico* ou *atributo qualitativo* assumido em determinado momento pela variável.

Exemplo:

Nome	Idade	Peso	Profissão	Nº de irmãos
Antônio	21	69	Padeiro	4
Manoel	42	86	Engenheiro	2
Maria	72	79	Farmacêutica	0

Alguns conjuntos de dados, como peso, consistem em números, enquanto outros são não numéricos, como profissão.

Aplicam-se as expressões *dados quantitativos* e *dados qualitativos* para distinguir os dois tipos:

a) Dados qualitativos (ou atributos): são separados em diferentes categorias por alguma característica *não-numérica*.

Ex:- sexo: masculino, feminino

- a cor da pele: branca, amarela, parda, preta.

b) Dados quantitativos: consistem em números que representam *contagens* ou *medidas*.

Ex: Salário dos operários em uma empresa.

Podemos ainda descrever os dados quantitativos distinguindo entre tipos *discretos* e *contínuos:*

a) Dados contínuos (numéricos): resultam de um número infinito de valores possíveis que podem ser associados a pontos em uma *escala contínua* de tal modo que não haja lacunas ou interrupções. Observa-se que a limitação prática de expressarmos qualquer valor dentro de uma escala contínua refere-se à precisão dos instrumentos de medida.

Ex.temperaturas mínimas obtidas em cada dia, quantidade de leite produzido em uma fazenda.

b) Dados discretos: resultam de um conjunto finito de valores possíveis, ou de um conjunto enumerável desses valores.

Ex.: conjunto de preços de um determinado produto, número de ovos produzidos em uma granja.

- **Dados brutos:** São dados que ainda não foram numericamente organizados.

Ex: Notas em uma prova de matemática:

4,5	5,0	8,5	4,0	10,0
5,5	7,0	7,0	4,5	5,7
6,3	9,2	8,0	5,0	6,5
6,2	4,9	8,5	9,0	5,0

- **Rol:** Conjunto de dados numéricos organizados em ordem crescente ou decrescente.

Ex: Notas em uma prova de matemática em ordem crescente:

4,0	4,0	4,5	4,9	5,0
5,0	5,0	5,5	5,7	6,2
6,3	6,5	7,0	7,0	8,0
8,5	8,5	9,0	9,2	10,0

Normalmente, procura-se evitar trabalhar com dados brutos, pois dificulta as observações e conclusões.

EXEMPLO PARA REFLEXÃO

O dono de uma loja de sapatos resolvecoletar dados sobre sua clientela para orientar suas vendas. Durante um período de 3 dias, ele anota em um caderno o número do tamanho de cadapar de sapato que é vendidos. Após este período ele verifica a seguinte anotação:

40 42 42 43 39 37 37 38 40 45 41 42 42 43 39 48 44 41 2 41 42 42

A fim de organizar os dados, qual seria o rol em ordem crescente dos dados brutos obtidos?

Após organizar os dados, é fácil responder algumas perguntas:

Qual o maior número vendido? Qual o menor? A maioria dos pares vendidos está acima ou abaixo de 40 ?

1.5 MÉTODOS ESTATÍSTICOS

Didaticamente, podemos considerar os seguintes passos para a aplicação do método estatístico na solução de um problema. Não se pode perder de vista, entretanto, que o aspecto essencial da Estatística é de proporcionar métodos inferenciais que permitam conclusões que transcendem os dados obtidos inicialmente. Assim, a análise e a interpretação dos dados estatísticos tornam possível o diagnóstico de uma empresa.

1.Definição do problema – devemos delinear cuidadosamente a questão a ser estudada. Certificar-se de que é clara a finalidade do estudo ou análise. Nesta fase inicial,o objetivo da pesquisa deve ficar definido.

2. Formulação de um plano para a coleta de dados – nesta fase devemos verificar qual é a população alvo, verificar como será feita a amostragem, qual a sua área de atuação, dimensionar o tamanho da amostra, como se fará a pesquisa, seleção do pessoal de campo, treinamento, recursos para desenvolvimento das atividades, dentre outros.

3. Coleta de dados estatísticos – são expressões numéricas que os fatos fornecem na área investigada.

Esta coleta pode ser direta ou indireta:

- Coleta Direta: é feita sobre elementos informativos de registro obrigatório ou quando os dados são obtidos pelo próprio pesquisador através de questionários.
  Ex. número de casamentos, número de divórcios, censo demográfico, etc.

- Coleta Indireta: ocorre quando os dados são obtidos através de elementos conhecidos.
  Ex.: A pesquisa sobre o número de divórcios no Brasil pode ser feita com os dados colhidos através da coleta direta obtidos pelos Cartórios.

Quanto ao tempo, a coleta pode ser classificada:

- **Periódica:** quando é feita em intervalos de tempo. Ex.: os censos.
- **Contínua:** quando realizada permanentemente. Ex.: os óbitos.
- **Ocasional:** quando efetuada sem época preestabelecida. Ex.: as epidemias

4. Crítica e apuração dos dados – a crítica tem por objetivo detectar possíveis falhas e imperfeições ocorridos na fase de coleta de dados ou durante o processamento desses dados para que não se incorra em erros que influenciem os resultados. A apuração dos dados é a soma e o processamento dos dados obtidos, além da disposição mediante os critérios de classificação.

5. Apresentação dos dados – os dados podem ser apresentados através de tabelas ou gráficos com a finalidade de tornar mais fácil o exame do que foi obtido.

6. Análise e interpretação dos resultados – é a tarefa fundamental do profissional que trabalha com Estatística. Elas podem partir do simples exame de tabela ou gráfico, ou podem utilizar métodos sofisticados de computação e emprego da matemática superior. Através da análise são obtidos dados importantes para conclusões como médias, tendências e dispersões. Através da análise de uma amostra significativa de uma população, é possível chegar a importantes conclusões sobre a população através dos métodos da Inferência Estatística.

1.6 AMOSTRAGEM:

É o processo de colher amostras. Nesse processo, cada elemento da população passa a ter a mesma chance de ser escolhido. É a parte inicial de qualquer estudo estatístico e consiste na escolha criteriosa dos elementos a serem submetidos ao estudo.

Há dois tipos de amostragens: probabilística e não probabilística. Para que possamos fazer inferências válidassobre a população apartir de uma amostra, é preciso que esta seja *representativa.*

Se você pensou em algo parecido como uma amostra que represente uma determinada população, com tendências as mais próximas possíveis, está correto. A amostra deve corresponder ao padrão da população. Imagine o resultado de uma pesquisa sobre quem foi o maior jogador de futebol do século em uma amostra com 400 argentinos e 20 brasileiros. Qual seria o resultado? Esta amostra estaria representando a opinião mundial?

Com certeza não, pois a amostra não foi selecionada de maneira aleatória. Quando isso acontece, dizemos que foi uma amostra *viciada.* Para que uma amostra seja representativa é preciso que seja de alguma forma aleatória.

Pense em uma moeda. Se você jogar a moeda para o alto saberá dizer se vai sair cara ou coroa? E um dado? Qual número sairá? Observe que não trata apenas de jogos. Imagine que você está em um banco e resolve verificar o sexo da primeira pessoa que entrar. Poderá dizer com antecedência se é homem ou mulher?

Fenômenos aleatórios: são aqueles que repetidos sob um mesmo conjunto de circunstâncias, nem sempre produzem os mesmos resultados.

A aleatorização é uma *condição indispensável* para a validação dos métodos empregados pela Estatística.

Os fenômenos aleatórios têm seguintes características:

- Podem ser repetidos sob as mesmas condições;
- Não se conhece, "a priori", um particular valor do resultado do experimento, embora possamos enumerar todos os resultados possíveis;
- Os resultados variam tornando impossível prever com exatidão o resultado do experimento em curso;
- Para um grande número de repetições os resultados tendem a apresentar proporções estáveis.

A grande dificuldade para um estatístico é obter amostras que sejam representativas da população. Isto é, a amostra deve reproduzir com perfeição a população.

Por que então não recorrer a um censo? Já que a amostra pode causar erros na estimação dos parâmetros populacionais, por que não utilizar o censo e obter diretamente o censo (quando toda a população é escolhida para estudo)?

Pense em uma pesquisa eleitoral. Por que não evitar o erro e utilizar um censo? Não seria melhor entrevistar todos os eleitores e perguntar em que candidato votaria?

Quem pensou que fazer um censo seria o mesmo que a eleição propriamente dita está no caminho certo, logo não compensaria realizar a pesquisa e seria melhor esperar a eleição.

Quando a população é muito grande ou infinita: o censo se torna impossível. O custo para entrevistar toda população brasileira é inimaginável. Imagine verificar qual a porcentagem de ovos de uma granja que estão estragados. Se for examinar todos os ovos (quebrando-os), no final você terminará sabendo ao certo quantos estão estragados, todavia como vender os ovos que você quebrou? Para amostras destrutivas o censo também é impossível.

Outro ponto a se considerar é o tempo gasto no censo. Considere que na eleição, quando os eleitores vão as urnas e não o contrário, utilizando a tecnologia da urna eletrônica, leva-se aproximadamente uma semana para computar todos os votos. Pense em quanto tempo levaria para entrevistar todos os eleitores.

A amostra pode ser mais atualizada que um censo: o tempo de realização do censo é consideravelmente maior do que a pesquisa amostral. No caso da pesquisa eleitoral, durante a realização do censo, as pessoas que já foram entrevistadas podem mudar de opinião em virtude de fatos novos. Por isso a pesquisa eleitoral é feita com amostras dimensionadas de modo que o número de elementos entrevistados seja o menor possível, em um menor tempo possível e repetida a intervalos pequenos de tempo, com a finalidadede atender o princípio da oportunidade.

Além dos dois principais fatores apresentados podemos acrescentar:

- o custo pode ser proibitivo;
- a precisão pode sofrer no caso de um censo de uma grande população. Uma amostra envolve um custo menor de observações e, consequentemente, menor número de coletores de dados, o que reduz o risco de erros; e
- a análise do tipo de informação pode definir o uso do censo ou de amostragem. Uma amostra, com o mesmo custo e mesmo tempo de um censo permite maior aprofundamento.

Por que então usar o censo?

- a população pode ser tão pequena que o custo e o tempo de um censo sejam um pouco maiores que uma amostra.
  Ex: 30 alunos de uma sala de aula

- quando o tempo da amostra é grande em relação ao da população e o esforço adicional requerido por um censo pode ser pequeno.
  Ex: se há grande variabilidade entre os itens de uma população, uma amostra deverá ser bastante grande para ser representativa.

- quando se exige precisão perfeita, então o censo é o único método aceitável.

CAPÍTULO 2

Distribuição de Frequências

2.1 DISTRIBUIÇÃO DE FREQUÊNCIAS

As Distribuições de Frequências – DF são séries *heterógradas*, dado que neste caso, todos os elementos – época, local e fenômeno – são fixos. O fenômeno ou fator especificadose apresenta agora através de gradações, isto é, os dados referentes ao fenômeno que se está representando são reunidos de acordo com sua magnitude.

Também chamada de Tabela de Freqüências, a DF serve para resumir grandes massas de dados brutos em tabelas que listam os valores observados em correspondência com as suas repetições verificadas.

A DF é mais fácil de consultar que o rol e, além disto, pode apresentar os dados agrupados em classes de valores ou não.

Ao usarmos DF, estamos buscando agilizar a manipulação e análise dos dados coletados, embora haja uma perda em precisão.

2.1.1 TABELA PRIMITIVA

Como foi visto anteriormente, após a coleta, os dados originais ainda não se encontram prontos para a análise por não estarem numericamente organizados. São os dados brutos. Consideremos os dados a seguir, relativos a uma coleta das estaturas de 40 alunos daturma A1, da 8ª série, da Escola Y (dados fictícios).

Estatura de 40 alunos – Turma A1 – 8ª Série – Escola Y
(em centímetros).

166	160	161	150	162	160	165	167	164	160
162	161	168	163	156	173	160	155	164	168
155	152	163	160	155	155	169	151	170	164
154	161	156	172	153	153	156	158	158	161

Os dados, assim dispostos, conforme foram coletados, sem a preocupação de organização, são os **dados brutos** e a tabela que os congrega, **tabela primitiva.**

Conhecidos os valores de uma variável, é difícil formarmos uma idéia exata do seu comportamento a partir de dados não ordenados. A ordenação pode ser crescente ou decrescente, porém a mais comum é a crescente. Conforme visto anteriormente, trata-se do rol. Assim, o rol abaixo nada mais é do que os dados brutos das 40 estaturas colocadas em ordem crescente de valores.

Estatura de 40 alunos
Turma A1 – 8ª Série – Escola Y (em centímetros).

150	151	152	153	154	155	155	155	155	156
156	156	157	158	158	160	160	160	160	160
161	161	161	161	162	162	163	163	164	164
164	165	166	167	168	168	169	170	172	173

A solução mais razoável é agruparmos os dados por faixas de valores da variável. Além disso, a variável estatura é uma variável contínua que se adapta perfeitamente a subdivisão em intervalos. Suponha que se deseja estabelecer intervalos constantes de 4 cm. Deste modo nossa tabela se apresenta do seguinte modo:

Estatura de 40 alunos
Turma A1 – 8ª Série – Escola Y

Estaturas (cm)	Frequencia
150 /--- 154	4
154 /--- 158	9
158 /--- 162	11
162 /--- 166	8
166 /--- 170	5
170 /--- 174	3
Total	40

O símbolo /--- significafechado a esquerda e aberto a direita. Assim, 150 /--- 154, significa em termos matemáticos, $150 \leq X < 154$. Logo, os quatro estudantes com estaturas neste intervalo são estudantes com estaturas 150,151,152 e 153 cm. Não é necessário iniciar nem terminar com os valores exatos que temos no rol. Ou seja, se o menor valor do rol fosse 151 cm poderíamos, mesmo assim, iniciar o intervalo com 150. O maior valor do rol é 173 cm. Entretanto, fechamos o último intervalo com 174. Poderíamos, em uma distribuição de notas, por exemplo, fazer os intervalos 0 /--- 2, 2 /--- 4, 4 /---6, 6 /--- 8 e 8 /--- 10. Se tivéssemos o valor 10 no rol, este seria incluído no último intervalo, pois por conversão sabe-se que o valor 10 está presente na contagem que define a última frequência.

Exercício

1. A partir do rol apresentado, organize uma distribuição de frequências considerando os intervalos sugeridos (0 /--- 2, 2 /--- 4, 4 /---6, 6 /--- 8 e 8 /--- 10).

4,0	4,0	4,5	4,9	5,0
5,0	5,0	5,5	5,7	6,2
6,3	6,5	7,0	7,0	8,0
8,5	8,5	9,0	9,2	10,0

2.1.2 HIPÓTESE BÁSICA DA TABULAÇÃO – HBT

A Hipótese Básica da Tabulação é, na realidade, uma convenção adotada de modo a padronizar a interpretação e o emprego de tabelas para eventuais cálculos e obtenção de medidas características dos fenômenos.

HBT: “A variável dentro da classe distribui-se linearmente. Supõe-se que todos os valores incluídos dentro dos limites de uma classe distribuem-se igualmente por todo o intervalo e, portanto, o valor médio de todos os dados de cada classe coincide, para fins de cálculos, com o ponto médio da classe.”

2.1.3 ELEMENTOS DE UMA TABELA DE DISTRIBUIÇÃO DE FREQUÊNCIA

a) Classe de Frequência: é um agrupamento conveniente de dados observados, visando facilitar o estudo do fenômeno. Classe é, portanto, cada um dos grupos de valores que constituem intervalos dentro dos quais estão as observações. A classe é identificada por seus extremos (**li** e **Li**) ou pela ordem que se apresenta na DF. O número da classe da DF nas fórmulas que serão apresentadas é identificado genericamente pelo índice **i**.

$$i = 1,2,3,...,n$$

b) Limites de Classe: são os extremos de cada classe: *Limite*inferior e limite superior de cada classe.

LI – Limite Inferior

LS – Limite Superior

c) Intervalo Total ou Amplitude Total (AT): é a diferença entre o limite superior da última classe (limite superior máximo) e o limite inferior da primeira classe (limite inferior mínimo).

$$AT = Lmax - L\ min$$

No exemplo das alturas da turma teremos:

L(máx) = 174 cm
L(mín) = 150 cm
AT = 174 – 150 = 24 cm

d) Amplitude Amostral (AA): é a diferença entre o valor máximo e o valor mínimo da amostra (ou população).

$$AA = X\ máx - X\ mín$$

No exemplo

x(máx) = 173 cm
x (mín) = 150 cm

AA = 173 – 150 = 23 cm

Normalmente a AT não coincide com a AA por razões de aproximação

e) Amplitude de um Intervalo de Classe: é a medida do intervalo que define a classe. É o tamanho da classe, ou seja, o intervalo entre o limite superior e o inferior da classe.

$$Hi = Li - li$$

Como no exemplo todos os intervalos são iguais (intervalos constantes), escrevemos simplesmente:

$$h = 154 - 150 = 4\ cm$$

f) Ponto Médio de uma Classe(xi): É o ponto que, segundo a Hipótese Básica da Tabulação, representa a classe, fazendo com que, para efeito de cálculo, todas as observações da classe sejam consideradas como tendo por valor x_i.

$$x_i = l_i + \frac{h_i}{2} = \frac{li + Li}{2}$$

No exemplo considerando a 1ª classe:

$$x_1 = li + \frac{hi}{2} = 150 + \frac{4}{2} = 152$$

g) Número de Classes (k): É a quantidade de classes existentes na DF.

Para DF com **h** constante temos que:

$$\mathbf{k} = \frac{\mathbf{AT}}{\mathbf{h}}$$

No exemplo:

$$k = \frac{24}{4} = 6$$

Um dos métodos existentes para sugerir o número de classes que deve possuir uma DF de h constante é através da relação empírica de Sturges, que é utilizado quando há falta de prática ou de conhecimento do fenômeno estudado e, consequentemente, não se sabe quantas classes a DF deverá ter. Geralmente o número de classes sugerido pela relação permite uma representação eficiente de dados. A relação é fornecida pela fórmula a seguir:

Relação de Sturges: $k = 1 + 3{,}3 \log n$

No exemplo:

$k = 1 + 3{,}3 \log 40 = 6{,}29 \cong 6$ (atendida a sugestão de Sturges)

Outro exemplo:

Notas da prova de matemática
Turma 701

Classes de Graus	Freqüência
0 ⊢— 5	4
5 ⊢— 6	9
6 ⊢— 8	12
8 ⊢— 10	5
Σ	30

2.1.4 MONTAGEM DE UMA DF COM CLASSES DE INTERVALOS IGUAIS

A montagem de DF com classes iguais é freqüentemente empregada quando não existe nenhuma razão que justifique a montagem de DF em classes diferentes, como no caso das notas dos alunos. Para padronizar a montagem de DF de classes iguais, deve ser seguido o seguinte roteiro:

1. Monte uma lista dos dados, de preferência um **ROL**;
2. Encontre AA, a amplitude total;
3. Estabeleça o valor de **k** ou o calcule pela relação de Sturges
   $$k = 1 + 3{,}3 \log n ;$$
4. Calcule $h = \frac{AA}{k}$; (**Obs: h sempre arredondado para cima**)
5. Determine os limites das classes;
6. Construa a DF, lembrando-se de incluir os elementos complementares.

Exercício:

A partir do rol apresentado, organize uma distribuição de frequências considerando os intervalos sugeridos (0 /--- 2, 2 /--- 4, 4 /---6, 6 /--- 8 e 8 /--- 10) seguindo o método proposto para DF a intervalos constante.

4,0	4,0	4,5	4,9	5,0
5,0	5,0	5,5	5,7	6,2
6,3	6,5	7,0	7,0	8,0
8,5	8,5	9,0	9,2	10,0

2.1.5 DISTRIBUIÇÃO DE FREQUÊNCIAS DE DADOS NÃO AGRUPADOS POR CLASSES DE VALORES

Em geral construímos este tipo de tabela DF quando a variável é discreta. Quando a variável é contínua normalmente observamos as frequências, dentro de intervalos regulares.

Número de televisores		Contagem	Frequência
0	II	2	
1	IIII	4	
2	IIIIIIII	8	
3	IIIIIIIIIII	11	
4	IIIIIIIIIII	11	
5	IIIIIIII	8	
6	III	3	
7	II	2	
8	I	1	
Total		50	

2.1.6 FREQUÊNCIAS

Frequência é o número de repetições do mesmo valor observado durante o levantamento estatístico. No caso de DF por classes, a frequência representa o número de observações dentro do intervalo.

As frequências são apresentadas na DF de diferentes formas dependendo do interesse do pesquisador. As frequências podem ser SIMPLES ou ACUMULADAS e ainda, dentro de cada situação, pode ser absoluta, relativa ou percentual.

a) Frequência Simples Absoluta - F_i

É o número de repetições de um determinado valor observado da VA. No caso de DF por classes é o número de observações dentro de um mesmo intervalo de classe. A soma das frequências simples é igual ao número total de observações **n**.

$$\sum_{i=1}^{m} F_i = n$$

b) Frequência Simples Relativa - f_i

É a proporção entre o número de observações de um valor individual ou de uma classe de valores e o número total de observações.

$$f_i = F_i / n$$

c) Frequência Simples Percentual - fp_i

É a frequência relativa expressa em termos percentuais.

$$f_{pi} = f_i \times 100$$

d) Frequência Acumulada Absoluta "Abaixo de" - $F_{a\,i}\uparrow$

É a soma das frequências simples absolutas da classe (ou valor) com as frequências simples absolutas das classes (ou valores) **anteriores**.

Usada toda vez que se procura saber quantas observações existem **ATÉ** uma determinada classe (ou valor) **INCLUSIVE**.

$$Fa_i\uparrow = Fa_{i-1}\uparrow + Fi$$

e) Frequência Acumulada Relativa "Abaixo de" - $f_{a\,i}\uparrow$ ou fad_i

$$fa_i\uparrow = fa_{i-1}\uparrow + fi$$

f) Frequência Acumulada Percentual "Abaixo de" - $fp_{a\,i}\uparrow$

$$fpa_i\uparrow = fpa_{i-1}\uparrow + fpi$$

g) Frequência Acumulada Absoluta "Acima de" - $F_{a\,i}\downarrow$

É a soma das freqüências simples absolutas da classe (ou valor) com as freqüências simples absolutas das classes (ou valores) **posteriores**.

Usada toda vez que se procura saber quantas observações existem **além** de uma determinada classe (ou valor) **inclusive**.

$$Fa_i\downarrow = Fa_{i-1}\downarrow - F_{i-1} \quad \text{ou} \quad Fa_{i-1}\downarrow = Fa_i\downarrow + F_{i-1}$$

RELAÇÃO IMPORTANTE

$$F_{a_i}\uparrow + F_{a_{i+1}}\downarrow = n$$

h) Frequência Acumulada Relativa "Acima de" - $f_{a_i}\downarrow$

$$f a_{i-1}\downarrow = f a_i \downarrow + f_{i-1}$$

i) Frequência Acumulada Absoluta "Acima de" - $f p_{a_i}\downarrow$

$$f p a_{i-1}\downarrow = f p a_i \downarrow + f p_{i-1}$$

Exemplo:

Número de livros lidos por 25 alunos em 1996

Valor	Fi	fi	fpi	Fai↑	Fai↓
0	5	0,20	20	5	25
1	4	0,16	16	9	20
2	15	0,60	60	24	16
3	1	0,04	4	25	1
Σ	25	1	100	---	---

Exercício

Complete a tabela abaixo:

Estaturas	Fi	xi	fi	Fa_i	Fad_i	fp_i
150 /--- 154	4					
154 /--- 158	9					
158 /--- 162	11					
162 /--- 166	8					
166 /--- 170	5					
170 /--- 174	3					
Total	40					

2.1.7 ARREDONDAMENTO DE DADOS

Para efeitos de padronização, a Cadeira de Estatística adota critérios a serem respeitados para fins de aproximação. Tais critérios servem para a montagem de tabelas e devem ser estendidosaos demais assuntos naquilo que for pertinente.

➡ As freqüências relativas serão apresentadas com quatro algarismos significativos após a vírgula e, para tanto, elas devem ser calculadas com cinco algarismos, sendo que o quinto será abandonado dentro da seguinte condição:

- Se maior ou igual a 5, o quarto algarismo é aproximado para cima.

  Ex: 0,07368 ≈ 0,0737

- Se menor que 5, apenas o abandonamos.

  Ex: 0,07324 ≈ 0,0732

➡ Cada coluna da tabela segue o número de casas decimais usadas para o valor da coluna que tiver maior número de casas não nulas, até o limite de quatro casas decimais.

➡ A soma das freqüências relativas e percentuais quando apresentarem valores diferentes de 1 e 100 respectivamente, tanto por falta como por excesso, será procedido o arredondamento na freqüência ou freqüências que levarem a um menor erro relativo, de modo que a soma seja exata. Para atender esta orientação, a Cadeira adota a correção na classe de maior **Fi**.

Em verificações, os cálculos intermediários serão feitos com no mínimo 4 casas decimais.

Exercícios

1. Os dados abaixo representam 15 observações de índices pluviométricosda cidade de Resende em mm de Hg obtidas no mês de Janeiro2017, pela EMPRESA DE ÁGUAS DO MUNICÍPIO.

144	152	159	160	160
151	157	146	154	145
141	142	143	158	141

2. Monte uma DF de dados agrupados em classes iguais e calcule as freqüências simples absolutas, relativas e percentuais, bem como as acumuladas absolutas crescentes e decrescentes.

Complete o extrato de DF abaixo:

Pesos (kg)	Fi	Fai↑	Fai↓
40├─50		5	
50├─60		15	75
60├─70			
70├─80			45
80├─90	30		
Σ			

CAPÍTULO 3

Representação Gráfica

3.1 CONCEITOS INICIAIS:

Já aprendemos a organizar os dados através de tabelas de distribuição de frequências. Na sequência natural dos trabalhos estatísticos que visam descrever o fenômeno, obtemos tabelas de distribuição de frequências ou séries estatísticas que são úteis para apresentar os resultados organizados e servem de base para a confecção dos respectivos gráficos.

Portanto, uma outra maneira de apresentar os resultados é através dos gráficos estatísticos. Os gráficos são importantes em qualquer apresentação, pois permitem uma visualização rápida, possibilitando, na maioria das vezes, um entendimento imediato do fenômeno em estudo.

Hoje, em cada mídia, seja em revistas especializadas, telejornais e internet, presenciamos uma verdadeira evolução gráfica dos diversos assuntos em pauta. E analisando cada aspecto temos que ter, pelo menos, a capacidade de interpretá-los.

A representação gráfica das séries estatísticas tem por finalidade permitir a visualização dos resultados, permitindo checar as conclusões sobre a evolução do fenômeno ou sobre como se relacionam os valores da série. Não há apenas uma maneira de representar graficamente uma série estatística. A escolha do gráfico apropriado ficará a critério do analista.

Contudo, os aspectos simplicidade, clareza e veracidade devem ser considerados na elaboração de um gráfico.

- FINALIDADES DOS GRÁFICOS:

Os gráficos possibilitam uma rápida visualização e compreensão do fenômeno; possibilitam a representação de distribuição de frequências e séries estatísticas, além depossibilitar a identificação de tendências associadas aos fenômenos.

- VANTAGENS:

Normalmente, os gráficos têm forte apelo visual, pois são figuras coloridas, muitas vezes associados a desenhos interessantes como caricaturas, por exemplo. São os chamados pictogramas. São, portanto, mais agradáveis de se observar e quase sempre são mais facilmente compreensíveis do que as tabelas. Além dessas vantagens, os gráficos são úteis para mostrar a tendência do fenômeno estudado.

- DESVANTAGENS:

Nem sempre são precisos epodem ser facilmente manipuláveis, no sentido de distorcer a informação (alterando-se a escala, por exemplo).

- CARACTERÍSTICAS DESEJÁVEIS:

→ simplicidade: não devem apresentar detalhes desnecessários
→ clareza: o gráfico deve esclarecer perfeitamente a informação que se deseja transmitir. Para isso,é preciso identificar as variáveis, apresentar o título, a fonte e outras informações julgadas necessárias.
→ veracidade: deve-se traduzir com exatidão a realidade dos fatos.

3.2 TIPOS DE GRÁFICOS ESTATÍSTICOS

Os Gráficos estatísticos são divididos em Gráficos de Variáveis Qualitativas, como:

- Gráficos de Barras.
- Gráficos de Setores.
- Gráficos Polares.

E Gráficos de Variáveis Quantitativas, como:

- Gráfico do tipo Histograma.
- Gráfico de Linhas.
- Ogiva ou Curva de Acumulação de Frequências.

3.2.1 GRÁFICOS EM COLUNAS OU BARRAS

É a representação de uma série por meio de retângulos, dispostos verticalmente (em colunas) ou horizontalmente (em barras). É um excelente gráfico para representarmos uma série histórica ou uma série geográfica. Cabe ressaltar que cada fenômeno a ser representado indica qual o gráfico mais adequado, ou seja, qual o gráfico que melhor e mais rapidamente transmite a informação que se deseja divulgar. Assim, pode ser que possamos utilizá-lo, mesmo para representar uma série especificativa.

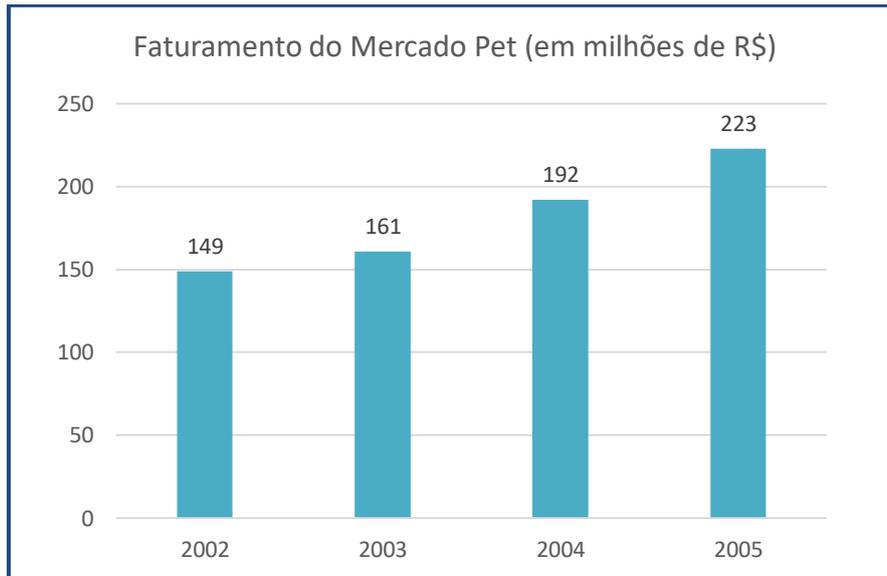



Gráfico 1: Faturamento do mercado *pet* entre 2002 e 2005
Fonte: Sindan/2007

Ex.: Veja a seguir, o gráfico de barras, usado para representar uma ***série geográfica.***

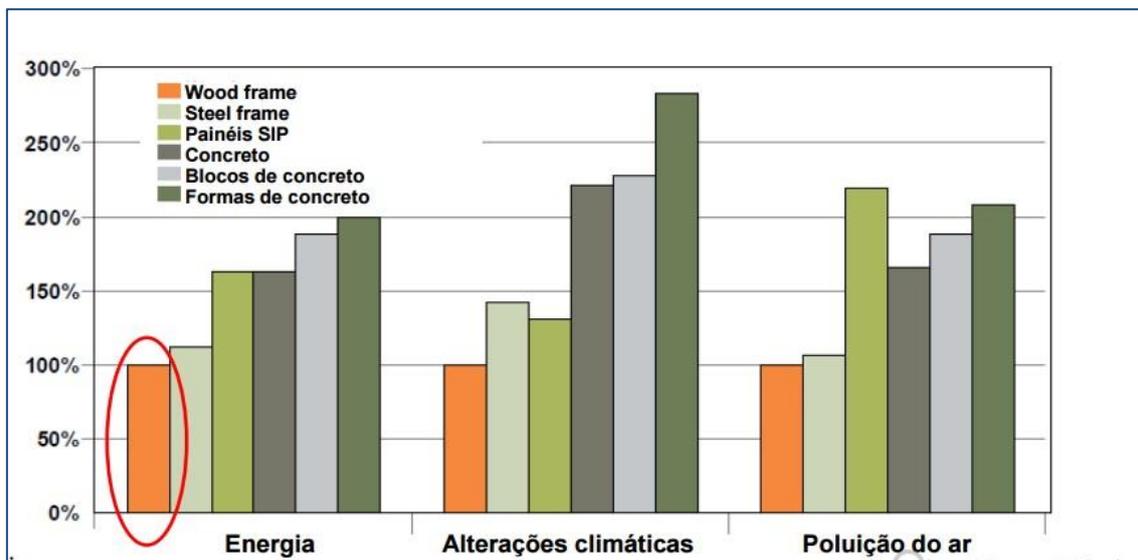



Gráfico 2: Países - representação de comparação do wood frame com outros serviços
Fonte: <www.aecweb.com.br> Acesso em: Set 2017

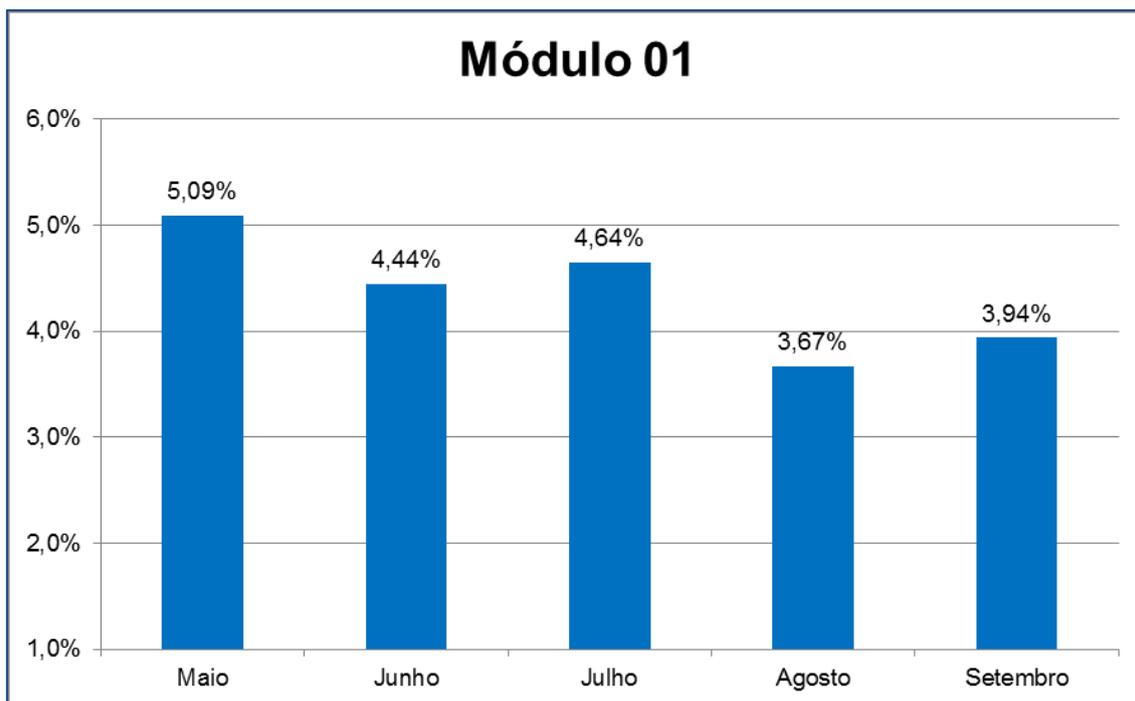



Gráfico 3: Redução do Desperdício Módulo 1.
Fonte: Autoria Própria

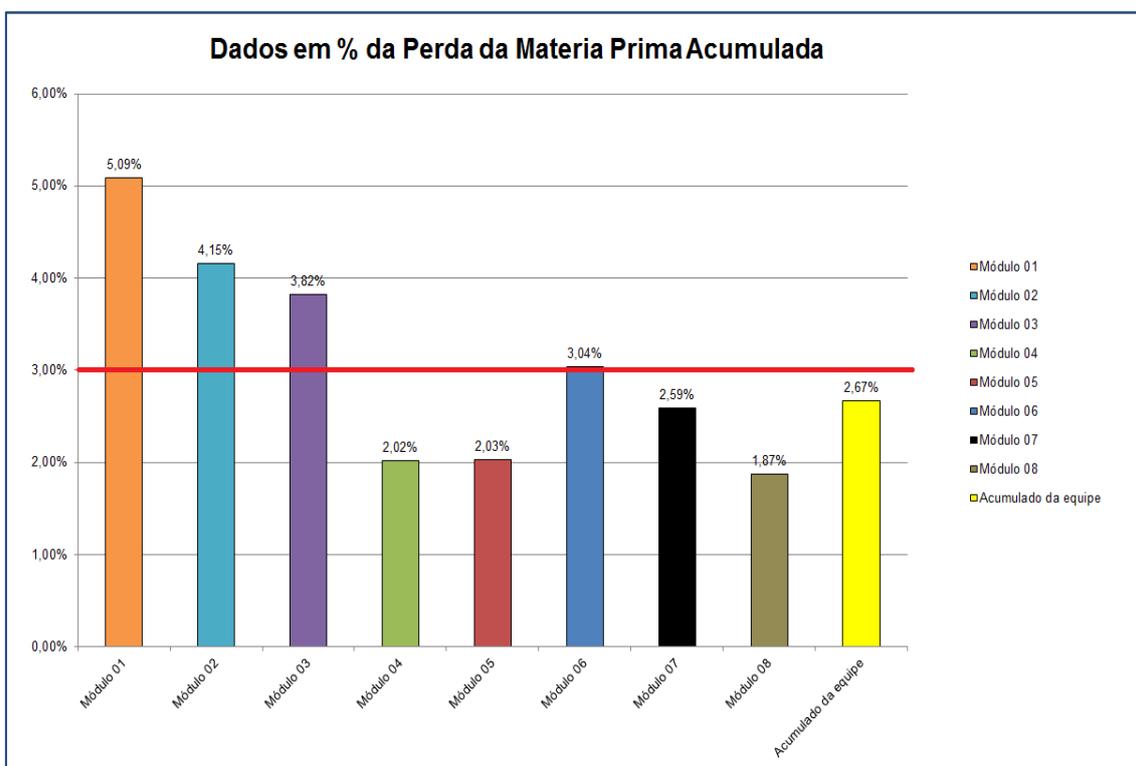



Gráfico 4 :Dados % Perda da Matéria-Prima Acumulada por Módulo
Fonte: Autoria Própria

3.2.2 GRÁFICO DE SETORES

Este gráfico é construído com base em um círculo e é empregado sempre que desejamos ressaltar a participação do dado no total.

O total é representado pelo círculo que fica dividido em tantos setores / partes.

Em geral, este tipo de gráfico pode representar bem todos os tipos de série. É claro que, dependendo da quantidade de setores que teremos que dividir o círculo, o gráfico pode tornar-se confuso. Assim, sempre é necessário avaliar que tipo de gráfico deve ser utilizado para a série em questão.

Exemplos:

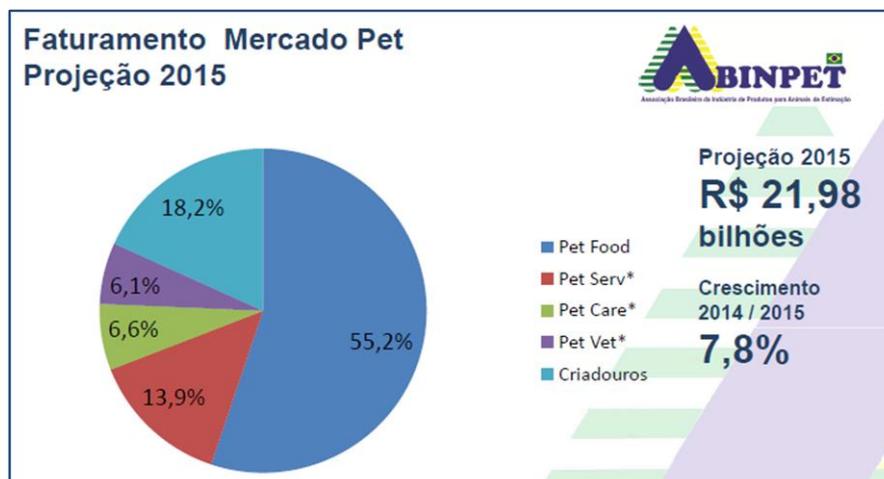



Gráfico 5: Representatividade de serviços do mercado *pet*
Fonte: Abinpet/2016

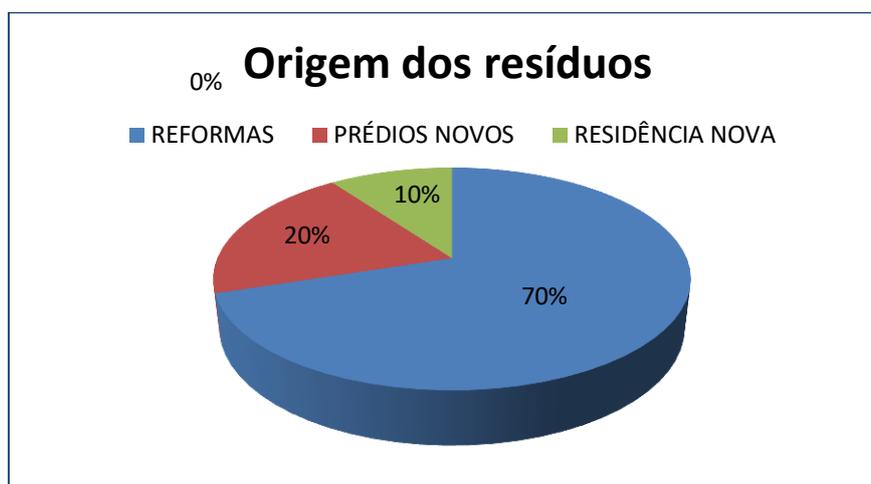



Gráfico 6 - Origem dos Resíduos
Fonte: MENEZES, 2011

3.2.3 GRÁFICOS DE LINHAS

Este tipo de gráfico se utiliza da linha poligonal para representar a série estatística.

É especialmente indicado para representar uma série histórica, pois mostra a tendência do fenômeno observado.

Exemplo:

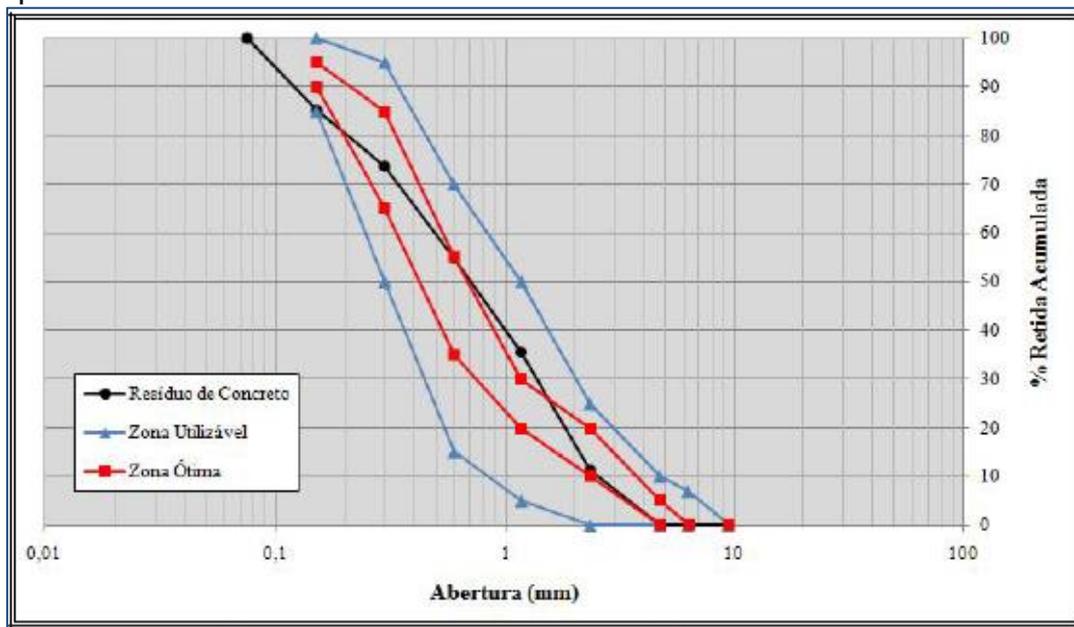



Gráfico 7 - Distribuição granulométrica do agregado miúdo artificial.
Fonte: SIMIELE, 2010

3.2.4 GRÁFICOS POLARES

É o gráfico ideal para representar séries temporais cíclicas, isto é, séries temporais que apresentam em seu desenvolvimento determinada periodicidade como, por exemplo, a variação da precipitação pluviométrica ao longo do ano ou da temperatura ao longo do dia, a arrecadação de impostos de uma região ou o consumo de energia elétrica durante um determinado período. Um exemplo é um gráfico de radar.

Exemplo:

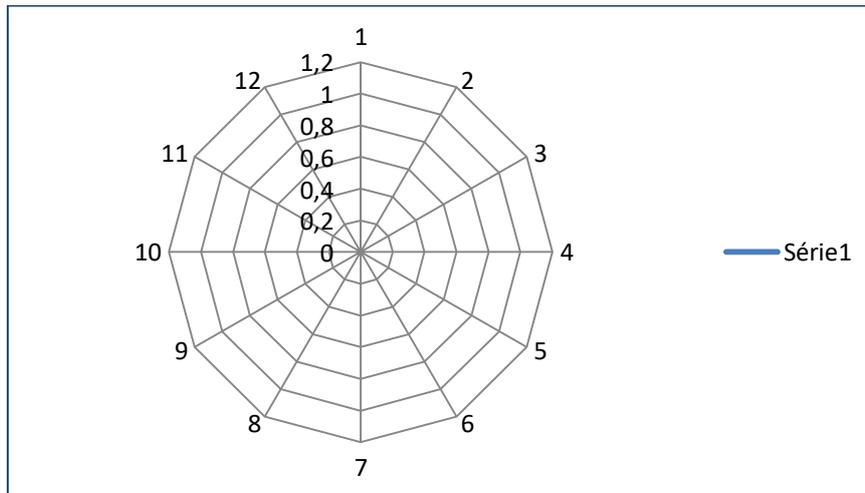



Gráfico 8– Gráfico de radar
Fonte: Próprio Autor

3.2.5 GRÁFICO HISTOGRAMA

O histograma, também conhecido como distribuição de freqüências, é a representação gráfica em colunas ou em barras (retângulos) de um conjunto de dados previamente tabulado e dividido em classes uniformes ou não uniformes.A base de cada retângulo representa uma classe. A altura de cada retângulo representa a quantidade ou a freqüência absoluta com que o valor da classe ocorre no conjunto de dados para classes uniformes ou a densidade de frequência para classes não uniformes. Importante ferramenta da estatística, o histograma também é uma das chamadas sete ferramentas da qualidade.

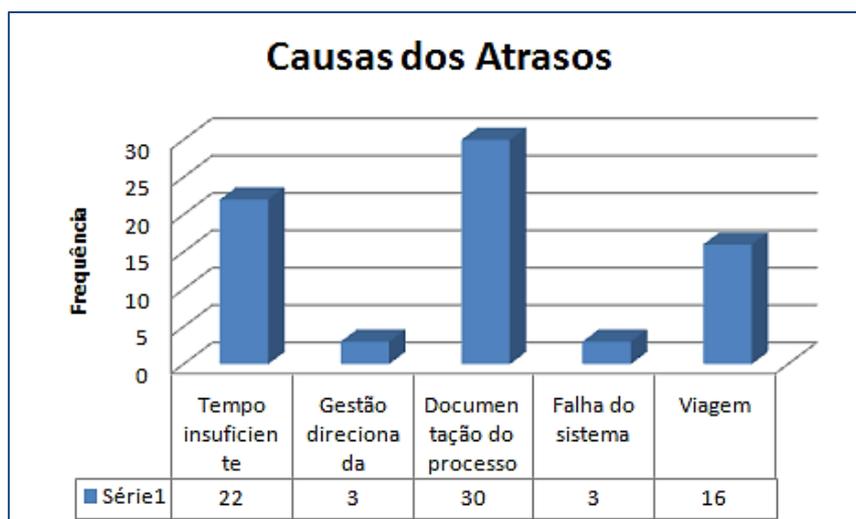



Gráfico 9– Exemplo de Gráfico de Histograma
Fonte:PMI (2013)

3.2.6 GRÁFICO OGIVA OU CURVA DE ACUMULAÇÂO DE FREQUÊNCIA

É um tipo de Gráfico de Linhas só que frequências acumuladas.

Pesos (kg)	Fi	Facu	Facu%
40|—50	10	10	24,4
50|—60	6	16	39
60|—70	15	31	75,5
70|—80	5	36	87,5
80|—90	5	41	100
Σ			

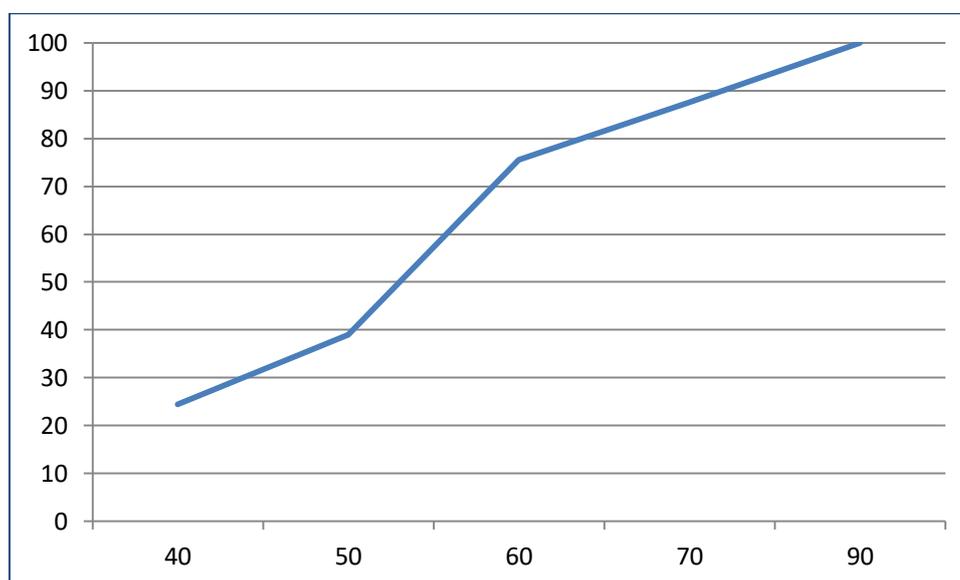



Gráfico 10 :Massa Corpórea dos 41 alunos de uma Turma
Fonte: Autoria Própria

3.2.7 CONCLUSÃO:

Dentre os gráficos apresentados, os de uso mais comum são: gráfico de linhas, de colunas e de setores.

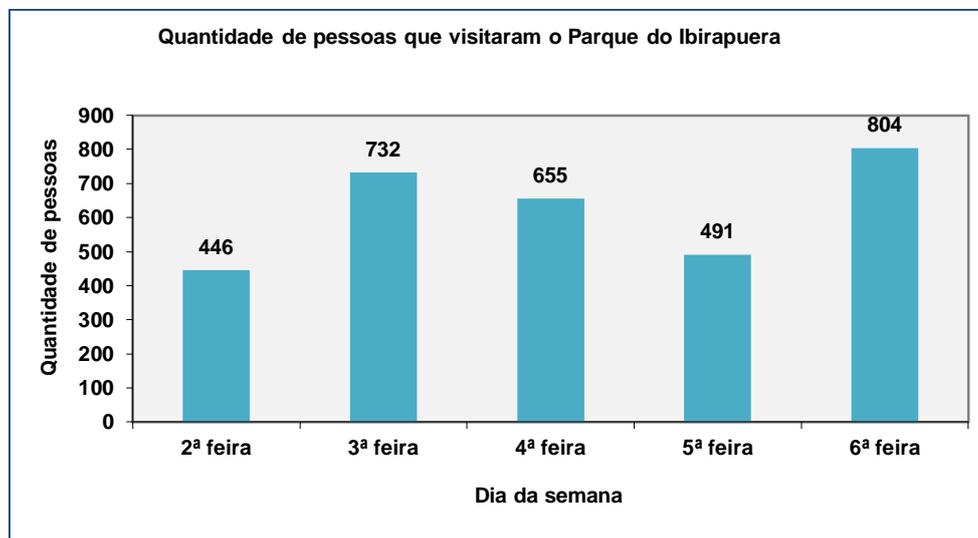
Ao observarmos um gráfico é preciso fazê-lo com um certo cuidado, no sentido de verificar se a sua construção está tecnicamente correta. Muitas vezes, a informação transmitida ou a impressão causada não corresponde à realidade dos dados.

É interessante observar que são incontáveis as possibilidades de apresentação dos gráficos. A variação de tipos, cores, apresentação em três dimensões,dentre vários outros aspectos, tudo isso aliado à facilidade e à rapidez de implementação nos aplicativos disponíveis nos programas de uso comum em computadores, trouxe uma contribuição importante à Ciência Estatística.

Não há regras e convenções rígidas que estabeleçam todos os detalhes construtivos. Como já falamos antes, o bom senso deve prevalecer. Desde a escolha do gráfico mais adequado, até os detalhes de apresentação como cores devem ser motivo de preocupação.

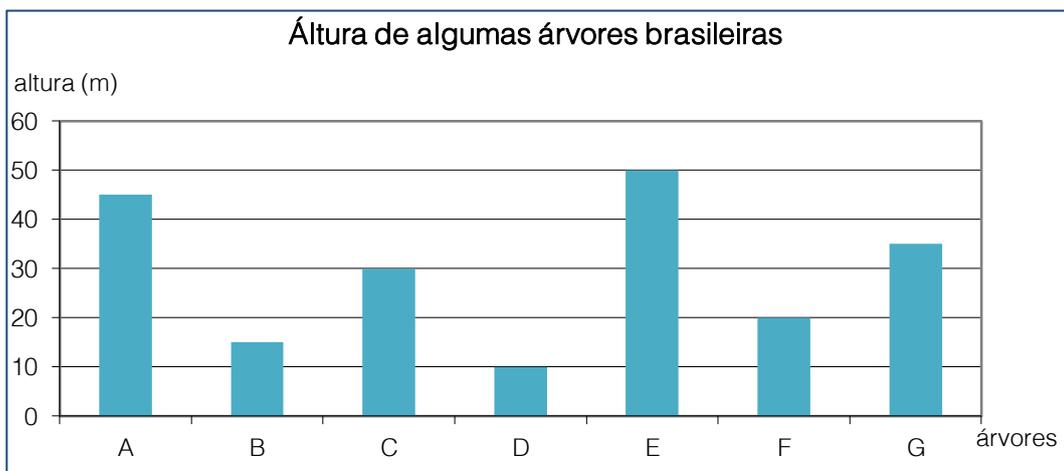
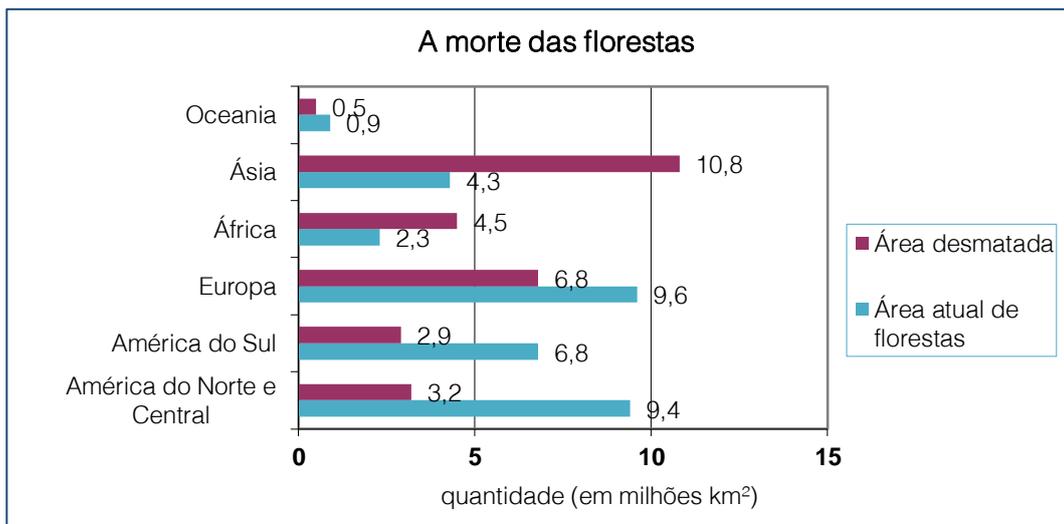
O gráfico histograma, bem como o conceito da curva de freqüências será importante para o entendimento de assuntos subseqüentes.

Exercícios

1) Observe o gráfico a seguir e responda:

a) Em qual dia da semana houve a maior quantidade de visitantes?
b) Em qual dia da semana houve menos visitantes?
c) Quantas pessoas, ao todo, visitaram o Parque do Ibirapuera nos cinco dias?

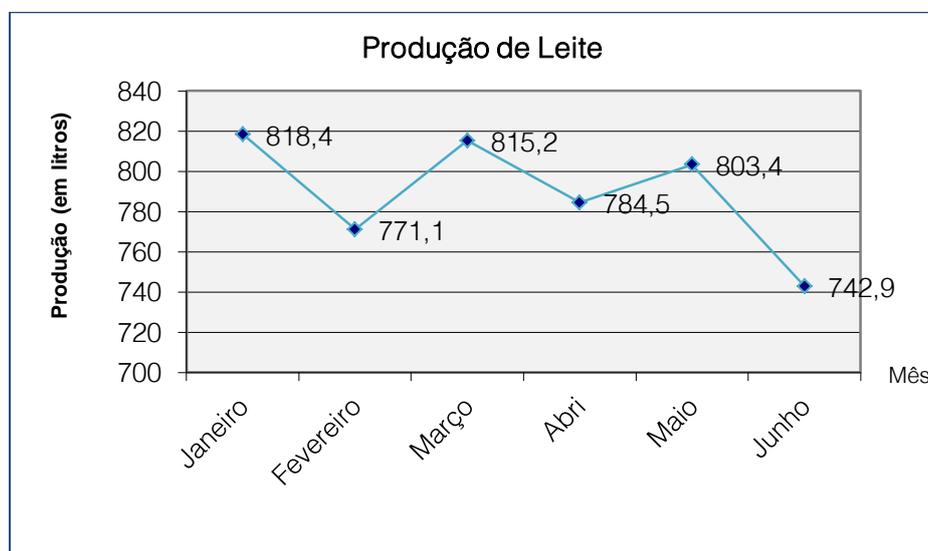
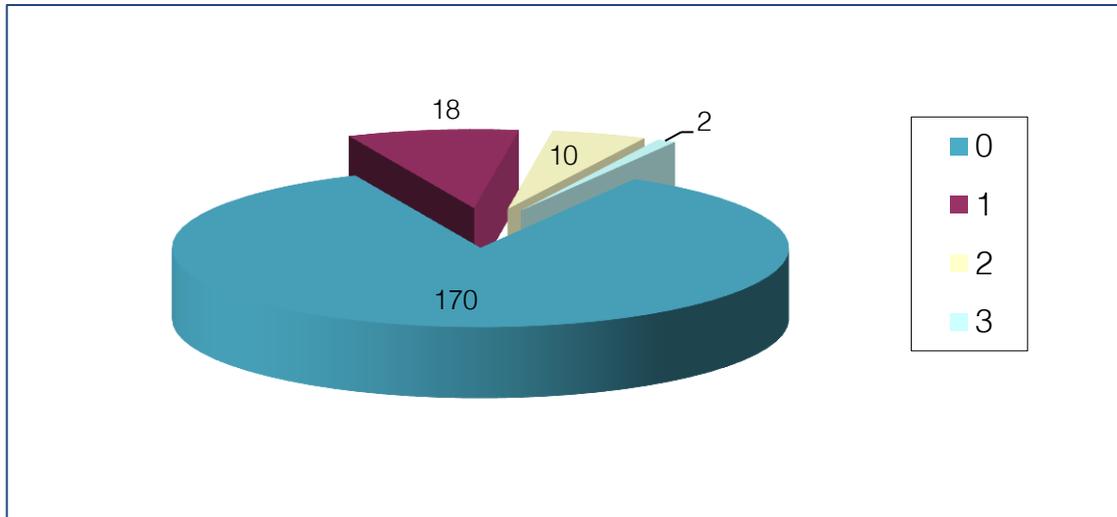
2) O estado das florestas do planeta e o que foi devastado pela ocupação humana são os dados que estão representados no gráfico a seguir. Observe estes dados que foram publicados na revista Época de 08/02/1999 e depois responda:

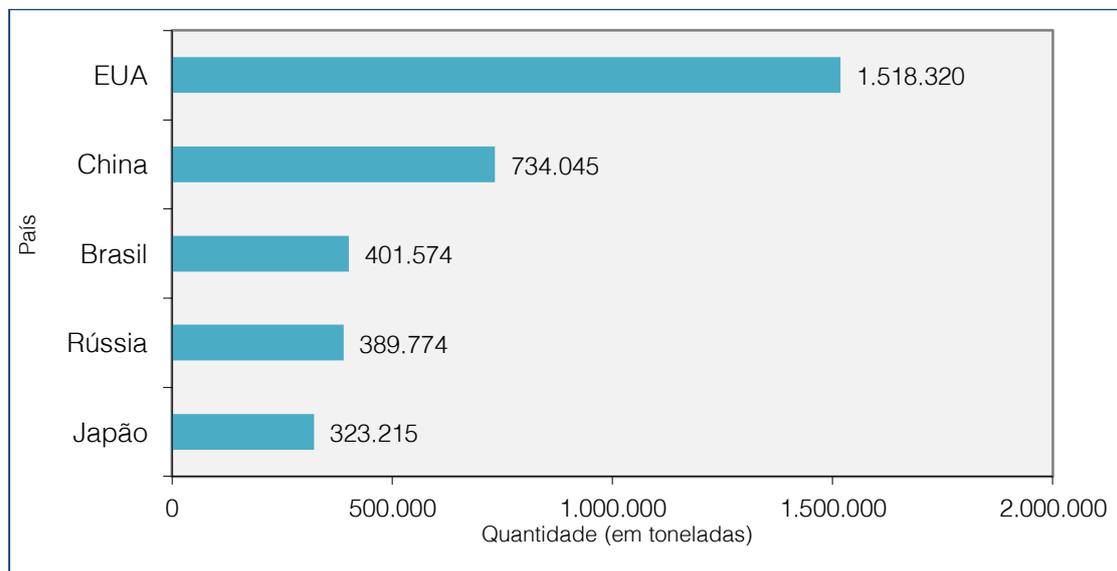
a) Em quais continentes mais da metade das florestas foi devastada pela ocupação humana?
b) Qual a área atual de florestas no mundo todo?
c) Qual a área desmatada no mundo todo?

3) O gráfico a seguir indica a altura máxima aproximada que algumas árvores brasileiras atingem.

De acordo com as informações apresentadas no gráfico e com os dados abaixo, identifique a árvore correspondente a cada coluna do gráfico e a altura máxima de cada árvore.

a) O jequitibá atinge 45 metros de altura.
b) O cedro atinge até 10 metros a menos que o jequitibá e 5 metros a mais que o pau-brasil.
c) O pau-brasil atinge 10 metros a mais que o abacateiro-do-mato e 14 metros a mais que a peroba.
d) A castanha-do-pará é cinco vezes maior que o cajueiro.

4) O gráfico de linhas abaixo mostra a produção de leite na Fazenda do Senhor BZerra no primeiro semestre do ano de 2006. Analise-o e responda:

a) Quantos litros de leite foram produzidos nesse semestre?
b) Quantos litros de leite foram produzidos, em média, por mês?
c) Quantos litros de leite, em média, foram produzidos diariamente no mês de janeiro?

5) O número de erros na primeira página de um jornal diário de grande circulação, em 200 dias pesquisados, está no gráfico a seguir:

Represente os dados (número de dias) em porcentagem.

6) Na revista Isto É de 23/02/2005 foi publicado a quantidade (em toneladas), dos países que mais emitiram CO_2 na atmosfera no ano de 2000.

Determine a diferença, em toneladas, de emissão de CO_2 entre:

a) EUA e Japão =
b) China e Rússia =
c) Brasil e Japão =
d) EUA e Rússia =

CAPÍTULO 4

Medidas de Tendência Central

4.1 CONCEITOS INICIAIS:

Em estatística, uma **tendência central** (ou, normalmente, uma **medida de tendência central**) é um valor central ou valor típico para uma distribuição de probabilidade. É chamada ocasionalmente como média ou apenas **centro** da distribuição. As medidas de tendência central mais comuns são a média aritmética, a mediana e a moda. Tendências centrais podem ser calculadas tanto para um número finito de valores quanto para uma distribuição teórica, a exemplo da distribuição normal. Ocasionalmente autores usam tendência central (ou **centralidade**) significando "a tendência de dados quantitativos de se agruparem ao redor de um valor central". Tal significado pode ser esperado da definição usual das palavras tendência e centralidade no dicionário. Autores podem julgar se dados têm tendência central forte ou fraca se baseando na dispersão estatística, medida pelo desvio padrão ou algo similar.

4.1.1MEDIDAS DE TENDÊNCIA CENTRAL:

Média para valores Discretos

a) Média Aritmética Simples: É o quociente entre somatório das variáveis pelo número de variáveis.

$$\bar{x}_a = \frac{\sum x_i}{n}$$

b) Média Aritmética Ponderada: É o quociente entre o produto dos valores de uma variável pelos seus respectivos pesos ou frequências pela soma dos pesos ou frequências.

$$\bar{x}_{ap} = \frac{\sum (f_i . x_i)}{\sum f_i}$$

c) Média Harmônica: É o inverso da média aritmética dos inversos das variáveis.

$$\bar{x}_h = \frac{n}{\sum \frac{1}{x_i}}$$

Ex.: Determinar a média harmônica de a e b.

$$\text{Mh} = \frac{2}{\frac{1}{2}+\frac{1}{8}} = \frac{2}{\frac{8+2}{8}} = \frac{2.8}{10} = 1{,}6$$

d) Média Geométrica: E a raiz n-ésima do produto de n valores.

$$\bar{x}_g = \sqrt[n]{x_1.x_2.x_3....x_n}$$

Exemplos:

1) A média geométrica dos números 2 e 8.

$$Mg(2{,}8) = \sqrt{2.8} = \sqrt{16} = 4$$

2) A média geométrica dos números 2,4 e 8.

$$Mg(2{,}4{,}8) = \sqrt[3]{2.4.8} = \sqrt[3]{64} = 4$$

Média para valores Agrupados

a) Média de Dados Agrupados:

- Sem intervalo de classe é o quociente entre o somatório dos produtos entre freqüências e variável pelos somatórios das frequências.

$$\overline{x} = \frac{\sum(f_i\, x_i)}{\sum f_i}$$

- Com intervalo de classe é o quociente entre o somatório do produto de freqüência e o ponto médio pelo somatório das freqüências simples.

$$\overline{x} = \frac{\sum(f_i\, X_i)}{\sum f_i}$$

Onde X_i éo ponto médio do intervalo de classe

Mediana

Uma segunda medida do meio de um conjunto de números é a **mediana (Md)**. Sua principal característica é dividir um conjunto *ordenado* de dados em dois grupos iguais; a metade terá valores *inferiores* à mediana, a outra metade terá valores *superiores* à mediana. Para calcular a mediana é necessário primeiro ordenar os valores do mais baixo para o mais alto. Em seguida, conta-se até a metade dos valores para achar a medida.

Mediana para valores Discretos

A mediana do conjunto 5,6 e 8 é 6, pois 6 está no meio.
Veja a seguinte lista de números:

8 2 6 5 5 6 4

Ordenando:

8 2 6 5 5 6 4

5 é a mediana do conjunto

Em geral, a mediana ocupa a posição:

$$\frac{(n+1)}{2}$$

No exemplo acima temos:

n = 7

$$\frac{(n+1)}{2} = \frac{7+1}{2} = 4$$

A mediana está na 4ª posição ⇒5
Veja outro exemplo:

7 8 9 10

$$\frac{(n+1)}{2} = \frac{4+1}{2} = 2{,}5$$

Como não há o 2,5º valor, utiliza-se a média entre o 2º e o 3º valores:

$$\frac{8+9}{2} = 8{,}5 \Rightarrow \text{Md} = 8{,}5$$

Em resumo, para se determinar a mediana:

1. Ordenar os valores.
2. Verificar se há um número ímpar ou par de valores.
3. Para um número ímpar a mediana é o valor do meio. Para um número par de valores, a mediana é a média dos dois valores do meio.

Observação: A escolha da média ou da mediana como medida de tendência central de um conjunto depende de diversos fatores. A média é sensível a cada valor do conjunto, inclusive os extremos. Por outro lado, a mediana é relativamente insensível aos valores extremos.

Considere os valores: 1,2,4, 6e 7. Tem-se

$$\bar{x} = \frac{1+2+4+6+7}{5} = 4$$

Md = 4 (posição central)

Suponha a seguinte substituição:

1 2 4 6 e 87

modificação no extremo

A mediana continuará igual a 4, pois o 4 continua como elemento central do conjunto. No entanto a média vai valer:

$$\bar{x} = \frac{1+2+4+6+87}{5} = 20$$

Assim, os dados sobre renda pessoal ou valor de casas de residências têm na mediana medida descritiva mais adequada; isso porque bastam alguns valores muito grandes para inflacionar a média aritmética.

De modo geral, a média possui certaspropriedades matemáticas que a tornam atraente. Além disso, a ordenação dos dados para determinar a mediana pode ser entediante e o cálculo da mediana não pode ser feito com máquina de calcular, ao contrário do que ocorre com a média.

Mediana para valores Agrupados

a) sem intervalos de classe. Dada a série:

nº de filhos (x_i) que se deseja ter	f_i	F_i
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
Total	34	

Então:

$$\frac{\sum f_i}{2} = \frac{34}{2} = 17$$

A menor freqüência acumulada que supera esse valor é 18, que corresponde ao valor 2 da variável.

Md = 2

No caso de $\frac{\sum f_i}{2} = F_i$ acontecer, a mediana será dada por: $Md = \frac{x_i + x_{i+1}}{2}$.

Exemplo:

i	nº de filhos (x_i) que se deseja ter	f_i	F_i
1	0	2	2
2	1	6	8
3	2	10	18
4	3	12	30
5	4	6	36
	Total	36	

$$\frac{\sum f_i}{2} = 18 = F_3 \text{, então: } Md = \frac{2+3}{2} = 2{,}5$$

Com Intervalos de Classe:

- Usa-se a Fórmula descrita abaixo:

$$Med = li + \left[\frac{\left(\frac{n}{2}\right) - fa_{[i-1]}}{fc} \right] . h$$

Onde:

- Li = Limite inferior da Classe que deve conter a Mediana.
- Fc = Frequência da Classe que deve conter a Mediana.
- Med = Mediana.
- Fa[i -1] = Frequência Acumulada da anterior da Classe que deve conter a Mediana.
- h = Tamanho da Classe.

Exemplo-1:

Classes (kg)	Fi	Fai
59 /--- 63	3	3
63 /--- 67	5	8
67 /--- 71	9	17
71 /--- 75	12	29
75 /--- 79	5	34
79 /--- 83	4	38
83 /--- 87	2	40
Total	40	

$$Med = 71 + \frac{(40/2 - 17).4}{12}$$

A mediana será a média entre os dois valores é Med = 72

Exemplo-2:

i	Total de pontos	f_i	F_i
1	150 I- 154	4	4
2	154 I- 158	9	13
3	158 I- 162	11	24
4	162 I- 166	8	32
5	166 I- 170	5	37
6	170 I- 174	3	40
	Total	40	

$\frac{\sum f_i}{2} = \frac{40}{2} = 20$, logo classe mediana é i = 3

λ = 158 F(ant) = 13 h = 4 f_3 = 11

$$Md = 158 + \frac{[20 - 13] \cdot 4}{11} = 158 + 2,5 = 160,5$$

No caso de $\frac{\sum f_i}{2} = F_i$ acontecer, a mediana será o limite superior da classe correspondente.

Moda

Uma terceira medida de tendência central é a Moda, a moda amostral de um conjunto de dados trata do valor que ocorre com maior frequência ou o valor mais comum em um conjunto de dados. Moda é especialmente útil quando os valores ou as observações não são numéricos, casos em que a média e a mediana não podem ser definidas. Por exemplo, a moda da amostra {Goiaba, Pêra, laranja, laranja, laranja, Uva} é laranja.Moda amostral não é necessariamente única como média ou mediana. Amostras que possuem uma moda são chamadas unimodais. Por exemplo, a amostra {1, 2, 4, 5, 5, 7, 9} tem moda 5. Amostras que possuem

duas modas são chamadas bimodais. Por exemplo, a amostra {1, 2, 3, 4, 4, 8, 8} tem modas 4 e 8. Amostras que possuem várias modas são chamadas multimodais. Por exemplo, a amostra {1, 2 3, 4, 4, 8, 8, 9, 9} tem modas 4, 8 e 9. Amostras que não tem moda são chamadas amodais. Por exemplo, a amostra {1, 3, 4, 8, 9, 10} não tem moda.

Moda para Valores Discretos

Basta verificar qual o valor que se repete
Exemplo: Na série { 1,5,7, 10, 10, 10, 12 , 13 } a moda é igual a 10.

Moda para Valores Agrupados

- Sem intervalo de classe: Somente pegar o valor de maior Frequência.

Exemplo:

Qual é a temperatura mais comum medida no mês de acordo com a tabela abaixo?

Temperatura(ºC)	Frequência
0º	3
10º	9
20º	12
30º	6

Resposta: 20º C é a temperatura modal, pois é a de maior frequência =12

- Com intervalo de classe: Fórmula de Czuber:

$$Mo = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

Onde:

- Limite inferior da Classe Modal = l.
- Diferença da freqüência da classe Modal e a Anterior $= \Delta 1$
- Diferença da freqüência da classe Modal e a Superior $= \Delta 2$
- Amplitude da Classe Modal = h.

Exemplo:

Calcular a Moda da Tabela abaixo:

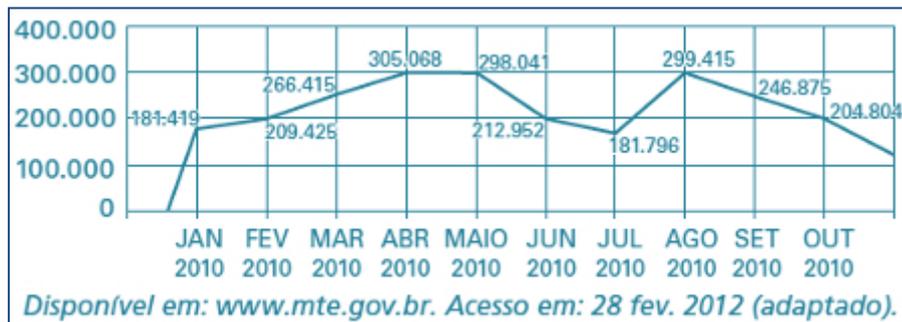
Classes (cm)	Frequência
54 I--------58	8
58 I--------62	11
62 I--------66	9
66 I--------70	5

$$Mo = 58 + 4\left(\frac{3}{3+2}\right) = 60{,}4 :$$

$$Mo = 60{,}4$$

Exercícios

1) (Enem 2012) O gráfico apresenta o comportamento de emprego formal surgido, segundo o CAGED, no período de janeiro de 2010 a outubro de 2010.

Disponível em: www.mte.gov.br. Acesso em: 28 fev. 2012 (adaptado).

Com base no gráfico, o valor da parte inteira da mediana dos empregos formais surgidos no período é:

a) 212.952 b) 229.913 c) 240.621 d) 255.496 e) 298.041

RESP: Letra b

2) A tabela a seguir mostra a evolução da receita bruta anual nos três últimos anos de cinco microempresas (ME) que se encontram à venda.

	2009 (em milhares de reais)	2010 (em milhares de reais)	2011 (em milhares de reais)
Alfinetes V	200	220	240
Balas W	200	230	200
Chocolates X	250	210	215
Pizzaria Y	230	230	230
Tecelagem Z	160	210	245

Um investidor deseja comprar duas das empresas listadas na tabela. Para tal, ele calcula a média da receita bruta anual dos últimos três anos (de 2009 até 2011) e escolhe as duas empresas de maior média anual.

As empresas que esse investidor decidiu comprar são:

a) Balas W e Pizzaria Y. b) Chocolates X e Tecelagem Z.

c) Pizzaria Y e Alfinetes V. d) Pizzaria Y e Chocolates X.

e) Tecelagem Z e Alfinetes V.

RESP: Letra d

3) Quais valores são, respectivamente, a moda, média e mediana dos números da lista a seguir?

133, 425, 244, 385, 236, 236, 328, 1000, 299, 325

a) 236; 361,1 e 312

b) 244; 361 e 312

c) 236; 360 e 312

d) 236; 361,1 e 310

e) 236; 361,1 e 299

RESP: Letra a

4) Dois alunos apostaram qual deles terminaria o ano com a maior média. As notas deles foram:

	1º Bimestre	2º Bimestre	3º Bimestre	4º Bimestre
Aluno 1	10,0	9,0	5,0	4,0
Aluno 2	6,0	6,5	7,5	8,0

Entre as alternativas a seguir, assinale aquela que for correta.

a) O aluno 1 conseguiu a melhor média, pois possui as melhores notas iniciais.

b) O aluno 2 conseguiu a melhor média, pois manteve as notas próximas umas das outras.

c) O aluno 1 venceu a aposta, pois sua média foi 7,0.

d) O aluno 2 venceu a aposta, pois sua média foi 7,0.

e) Nenhum aluno venceu a aposta, pois suas médias foram iguais.

RESP: Letra e

5) Comprei **5** doces a **R$ 1,80** cada um, **3** doces a **R$ 1,50** e **2** doces a **R$ 2,00** cada. O preço médio, por doce, foi de:

a) R$ 1,75 b) R$ 1,85 c) R$ 1,93 d) R$ 2,00 e) R$ 2,40

RESP: Letra a

6) Um professor de matemática costuma verificar a aprendizagem de seus alunos através da mediana das notas obtidas pela turma. Considere que a turma de 2014 obteve as seguintes notas no 2° bimestre:

2,5	7,6	8,9	9,1
3,4	6,7	5,8	6,9
10,0	8,6	5,7	4,9
8,3	6,7	7,9	9,2
8,1	7,5	8,3	1,9

Qual é a mediana das notas? Considerando que a média escolar é 7,0, a mediana está acima ou abaixo dessa média?

R: 7,55

7) Confira na tabela a seguir as medalhas conquistadas pelo Brasil nas Olimpíadas de 1968 a 2012:

Encontre a mediana do total de medalhas conquistadas pelo Brasil nesses anos.

Ano	Ouro	Prata	Bronze
1968	0	1	2
1972	0	0	2
1976	0	0	2
1980	2	0	2
1984	1	5	2
1988	1	2	3
1992	2	1	0
1996	3	3	9
2000	0	6	6
2004	5	2	3
2008	3	4	8
2012	3	5	9

RESP: 7 medalhas.

CAPÍTULO 5

Medidas de Dispersão

5.1 CONCEITOS INICIAIS:

Em Estatística, dispersão (também chamada de variabilidade ou espalhamento) mostra o quão esticada ou espremida é uma distribuição (teórica ou que define uma amostra). Exemplos comuns de medidas de dispersão estatística são a variância do desvio padrão e a amplitude interquartil. Dispersão é contrastada com posição ou tendência central e, juntas, elas são as propriedades de distribuições mais usadas.

5.2 MEDIDAS DE DISPERSÃO:

São necessários dois tipos de medidas para descrever adequadamente um conjunto de dados. Além da informação quanto ao “meio” de um conjunto de números, é conveniente ter também de um método que permita exprimir a dispersão. As medidas de dispersão indicam se os valores estão relativamente próximos uns dos outros, ou separados.Servem para verificar a representatividade das medidas de posição, pois é muito comum encontrar séries que, apesar de terem a mesma média, são compostas de maneira distinta.

Por exemplo, para as séries:

a) 20, 20, 20, 20, 20

b) 15, 10, 20, 25, 30

Tem-se $\bar{x}_a = \bar{x}_b = 20$

Notamos que os valores da série “a” se concentram totalmente na média 20, enquanto os valores da série “b” se dispersam em torno do mesmo valor. Ou seja, asérie “a” não apresenta dispersão e os valores da série “b” estão dispersos em torno de 20. Portanto,devemos medir o grau de concentração ou dispersão dos dados em torno da média. Para isso, calculamos as medidas de dispersão, que ora será descrita a seguir:

Amplitude Total

E a diferença entre o maior e o menor dos valores da série.É indicado por:

$$R \text{ ou } At = (X_{máx} - X_{mín}) \text{ ou } (L_i - l_i)$$

Exemplo:Para a série 10, 15, 22, 23, 25, 37, 38 $\Rightarrow$ R = 38 – 10 = 28

A utilização da amplitude total como medida de dispersão é muito limitada, pois, sendo uma medida que depende apenas dos valores externos, é instável, não sendo afetada pela dispersão dos valores internos.

A vantagem de utilizar o intervalo como medida de dispersão é de que o intervalo é relativamente fácil de calcular, mesmo para um grande conjunto de números. A significação do intervalo é fácil de entender. Infelizmente, algumas outras medidas de dispersão não são tão fáceis de entender, pelo menos intuitivamente.

Intervalo semi-interquartílico

Enquanto o desvio padrão e a variância calculam a medida de dispersão sem levar em conta a ordem dos dados, o intervalo interquartil avalia a dispersão de dados somente depois de ordená-los em ordem crescente. O intervalo interquartil é calculado com base no cálculo de quartis:o primeiro quartil (inferior), o quartil intermediário (mediana) e o terceiro quartil (superior) que estão ligados ao conceito de quartil. A diferença entre o quartil superior e o quartil inferior determina o intervalo interquartil.

É utilizado para verificar a dispersão em torno da mediana. É dado por:

$$I = \frac{Q_3 - Q_1}{2}$$

Exemplo: Foram calculadas as seguintes medidas para as notas dos alunos em duas matérias:

Física. $Q_1 = 3$; $Q_3 = 6{,}5$; $\bar{x} \cong 5$

Matemática: $Q_1=2$; $Q_3 = 7{,}0$; $\bar{x} \cong 5$

Observamos que, apesar de as duas distribuições apresentarem a mesma mediana, as notas de Física apresentam menor dispersão:

$$I_F = \frac{6{,}5 - 3}{2} = 1{,}75 \qquad \mathrm{I_M} = \frac{7-2}{2} = 2{,}5$$

Desvio Médio

Neste caso, considera-se o módulo de cada desvio $\left(x_i - \overline{\overline{x}}\right)$,evitando com isso que $\sum d_i = 0$. Assim, o Desvio Médio é dado por:

$$D_M = \frac{\sum \left|x_i - \overline{\overline{x}}\right|.F_i}{n} = \frac{\sum |d_i|.F_i}{n}$$

É a média aritmética dos desvios considerados em valores absolutos (módulos).

Variância

Considera-se o quadrado de cada desvio $\left(x_i - \overline{x}\right)^2$ evitando com isso que

$$\sum d_i = 0$$

Assim, a definição da variância é dada por:

$$\sigma^2 = \frac{\sum \left(x_i - \overline{\overline{x}}\right)^2 .f_i}{n} = \frac{\sum d_n^2 .f_i}{n}$$

É a média aritmética dos quadrados dos desvios.

Observação: σ^2 indica variância e lê-se sigma ao quadrado e $\overline{x}$ é a média da população.

Para o caso do cálculo da variância de valores amostrais é conveniente usarmos a seguinte fórmula:

$$S^2 = \frac{\sum \left(x_i - \overline{\overline{x}}\right)^2 .f_i}{n-1} = \frac{\sum d_i^2 .f_i}{n-1}$$

Observação: As diferenças entre as fórmulas são:

Para o caso de variância populacional (σ^2), usa-se a média populacional

($\bar{x}$), tendo como amostra (n).

Para o cálculo de variância amostral (S^2), usa-se a média amostral ($\bar{x}$), tendo como denominador o tamanho da mostra menos um (n-1).

Desvio Padrão

Observando a fórmula para o cálculo da variância, notamos tratar-se de uma soma de quadrados. Dessa forma, se a unidade da variável for, por exemplo, metro (m), teremos como resultado metro ao quadrado (m^2). Para voltarmos à variável original, necessitamos definir uma outra medida de dispersão, que é a raiz quadrada da variância – o desvio padrão. Assim:

$\sigma = \sqrt{\sigma^2}$ Desvio padrão populacional

$S = \sqrt{S^2}$ Devio padrão amostral

Desvio Padrão – Valores Discretos

Ou seja, para o cálculo do desvio padrão deve-se primeiramente determinar o valor da variância e, em seguida, calcular a raiz quadrada desse resultado.

Exemplo: Calcular o desvio médio, a variância e o desvio padrão da seguinte distribuição amostra:

x_i	f_i	$x_i.f_i$	$\lvert x_i - \bar{x} \rvert = \lvert d_i \rvert$	$\lvert d_i.f_1 \rvert$
5	2	10	3,06	6,12
7	3	21	1,06	3,18
8	5	40	0,06	0,3
9	4	36	0,94	3,76
11	2	22	2,94	5,88
Total	16	129		19,24

1º) Cálculo da média:

$$\bar{x} = \frac{\sum x_i . f_i}{n} = \frac{129}{16} = 8{,}06$$

2º) Cálculo do desvio médio:

$$D_M = \frac{\sum |x_i - \bar{x}| . f_i}{n} = \frac{\sum |d_i| . f_i}{n} = \frac{19{,}24}{16} = 1{,}2$$

3º) Cálculo da variância amostral:

Para isso temos que inserir mais uma coluna para: $\sum x_i^2 f_i$.

x_i	f_i	$x_i . f_i$	$\sum x_i^2 f_i$	$\lvert x_i - \bar{x} \rvert = \lvert d_i \rvert$	$\lvert d_i . f_1 \rvert$
5	2	10	50	3,06	6,12
7	3	21	147	1,06	3,18
8	5	40	320	0,06	0,3
9	4	36	324	0,94	3,76
11	2	22	242	2,94	5,88
Total	16	129	1083		19,24

$$S^2 = \frac{1}{n-1}\left[\sum x_i^2 . F - \frac{(\sum x_i . F_i)^2}{n}\right] = \frac{1}{16-1}\left[1083 - \frac{(129)^2}{16}\right] = \frac{1}{15}\left[\frac{17328 - 16641}{16}\right] = 2{,}86$$

4º) Cálculo do desvio padrão amostral:

$$S = \sqrt{S^2} = \sqrt{2{,}86} \cong 1{,}69$$

Conclusão: A distribuição possui média igual a 8,06. Ou seja, seus valores estão em torno de 8,06 e seu grau de concentração é de 1,2, medido pelo Desvio Médio e de 1,69 medido pelo Desvio padrão.

Desvio Padrão – Valores Agrupados

Exemplo: Calcular o Desvio Padrão da Tabela abaixo que mostra as notas de 43 Acadêmicos da Turma A.

Tabela-1: Notas da atividade 1, dos acadêmicos da Turma A

Notas	Acadêmicos	x_i	$x_i f_i$	$\lvert x_i - 6{,}51 \rvert$	$\lvert x_i - 6{,}51 \rvert^2 f_i$
3,8 ├ 4,8	5	4,3	21,5	2,21	24,4205
4,8 ├ 5,8	6	5,3	31,8	1,21	8,7846
5,8 ├ 6,8	12	6,3	75,6	0,21	0,5292
6,8 ├ 7,8	15	7,3	109,5	0,79	9,3615
7,8 ├ 8,8	5	8,3	41,5	1,79	16,0205
Total	43		279,9		59,1163

Para o cálculo do Desvio padrão precisamos da soma dos quadrados dos módulos dos desvios multiplicados pela frequência. Em seguida, é só aplicarmos a fórmula:

$$S = \sqrt{\frac{\sum_{j=1}^{k} \left(x_j - \overline{x}\right)^2 f_j}{n-1}} = \sqrt{\frac{\sum_{j=1}^{8} \left(x_j - 6{,}51\right)^2 f_j}{43-1}} = \sqrt{\frac{59{,}12}{42}} = \sqrt{1{,}41} \Rightarrow S = 1{,}19$$

e) Coeficiente de Variação (CV) de Pearson:

Esse coeficiente é dado pela razão entre o desvio-padrão e a média referentes aos dados de uma mesma série.

$$CV = \frac{\sigma}{\mu}\, ou\, \frac{s}{\overline{x}}$$

Como o coeficiente de variação analisa a dispersão em termos relativos, ele será dado em %. Quanto menor for o valor do coeficiente de variação, mais homogêneos serão os dados, ou seja, menor será a dispersão em torno da média. De uma forma geral, se o CV:

For menor ou igual a 15% → baixa dispersão: dados homogêneos

For entre 15 e 30% → média dispersão

For maior que 30% → alta dispersão: dados heterogêneos

Exemplo:

Em um grupo de moradores de uma determinada região foram analisadas a idade (em anos) e a altura (em metros) das pessoas. Deseja-se comparar a dispersão em termos relativos em torno da média dos dois conjuntos de dados, a fim de verificar qual deles é mais homogêneo. Na coleta dos dados verificou-se que:

Idade das pessoas: $\overline{x}$=41,6 e S = 0,82
Altura das pessoas: $\overline{x}$=1,67 e S = 0,2

Qual conjunto de dados apresenta menor dispersão relativa em torno da média?

Solução: O primeiro fato a se observar é que os dados analisados possuem unidades de medida diferentes. Dessa forma, somente o desvio padrão não é suficiente para comparar os dois conjuntos. Nesse caso, é preciso calcular o coeficiente de variação para fazer a comparação da variação em torno da média dos dados.

Assim, teremos:

Cálculo do CV da idade.

$$CV = \frac{s}{\overline{X}} \cdot 100$$

$$CV = \frac{0{,}82}{41{,}6} \cdot 100$$

$$CV = 0{,}0197 \cdot 100$$

$$CV = 1{,}97\%$$

Cálculo do CV da altura.

$$CV = \frac{s}{\overline{X}} \cdot 100$$

$$CV = \frac{0{,}2}{1{,}67} \cdot 100$$

$$CV = 0{,}119 \cdot 100$$

$$CV = 11{,}9\%$$

Interpretação dos dados: como o coeficiente de variação da idade foi menor que o coeficiente de variação da altura, pode-se afirmar que os dados relativos à idade são mais homogêneos que os dados da altura.

Exercícios

1) Calcular o Desvio padrão da tabela abaixo:

Tabela 4.13 Distribuição de freqüências do consumo de energia elétrica

Consumo	Número de usuários f_i	x_i	$x_i f_i$	$(x_i - 79{,}5)$	$(x_i - 79{,}5)^2$	$(x_i - 79{,}5)^2 f_i$
5 ├25	4	15	60	- 64,5	4160,25	16641,0
25 ├45	6	35	210	- 44,5	1980,25	11881,5
45 ├65	14	55	770	- 24,5	600,25	8403,5
65 ├85	26	75	1950	- 4,5	20,25	526,5
85 ├105	14	95	1330	15,5	240,25	3363,5
105 ├125	8	115	920	35,5	1260,25	10082,0
125 ├145	6	135	810	55,5	3080,25	18481,5
145 ├165	2	155	310	75,5	5700,25	11400,5
$\sum$			6360			80780

A média aritmética do consumo já foi calculada anteriormente:

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{6360}{80} = 79{,}5$$

Cálculo do Desvio-padrão pela Fórmula Original :

$$S = \sqrt{\frac{\sum_{j=1}^{k} (x_j - \bar{x})^2 f_j}{n-1}} = \sqrt{\frac{\sum_{j=1}^{8} (x_j - 79{,}5)^2 f_j}{80-1}} = \sqrt{\frac{80780}{79}} = \sqrt{1022{,}53} \Rightarrow S = 31{,}98$$

O desvio-padrão do consumo de energia elétrica é 31,98 Kwh. Lembre-se que o desvio médio já foi calculado, resultando em D_m = 24,63 Kwh.

2) Calcular o Desvio Padrão para os dados abaixo:

Posições	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Valores Ordenados	1	2	2	3	3	3	3	4	5	8	11	11	12	16	22	23	37	39	42	45	46	47	51	52	56	90

$$Var = S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2}{n-1} = \frac{(3^2 + \dots + 52^2) - 26(24{,}38)^2}{25} = 555{,}61$$

Assim, desvio padrão = $S = \sqrt{S^2} = \sqrt{555{,}61} = 23{,}57$

3) Imagine a seguinte situação: o dono de uma microempresa pretende saber, em média, quantos produtos são produzidos por cada funcionário em um dia. O chefe tem conhecimento que nem todos conseguem fazer a mesma quantidade de peças, mas pede que seus funcionários façam um registro de sua produção em uma semana de trabalho. Ao fim desse período, chegou-se à seguinte tabela:

	Quantidade de peças produzidas por dia				
Funcionários	Segunda	Terça	Quarta	Quinta	Sexta
A	10	9	11	12	8
B	15	12	16	10	11
C	11	10	8	11	12
D	8	12	15	9	11

Para saber a produção média de seus funcionários, o chefe faz o cálculo da média aritmética $(\bar{x})$ de produção, isto é, a soma do número de peças produzido em cada dia dividida pela quantidade analisada de dias.

A partir desse cálculo, temos a produção diária média de cada funcionário. Mas se observarmos bem a tabela, veremos que há valores distantes da média. O funcionário **B**, por exemplo, produz uma média de **12,8 peças por dia.** No entanto, houve um dia em que ele produziu **16 peças** e outro dia em que ele confeccionou apenas **10 peças**. Será que o processo utilizado pelo dono da empresa é suficiente para o seu propósito?

Para esse exemplo, ficou fácil concluir que há uma grande variação entre a produção de cada funcionário. Mas e se essa fosse uma grande empresa, com mais de mil funcionários, ou se fosse observada a produção em um ano, será que conseguiríamos definir essa variação com tanta facilidade?

O estudo da Estatística apresenta **medidas de dispersão** que permitem a análise **da dispersão dos dados**. Inicialmente veremos a **variância**, *uma medida de dispersão que mostra quão distantes os valores estão da média.* Nesse caso, como estamos analisando todos os valores de cada funcionário e não apenas uma "amostra", trata-se do cálculo da **variância populacional (var)**.

O cálculo da variância populacional é obtido através da soma dos quadrados da diferença entre cada valor e a média aritmética, dividida pela quantidade de elementos observados. Observe o cálculo simplificado para esse exemplo:

$$\text{var} = \frac{(\text{segunda} - \text{média aritmética})^2 + (\text{terça} - \text{média aritmética})^2 + \cdots + (\text{sexta} - \text{média aritmética})^2}{\text{quantidade de dias}}$$

Observação: Se estivéssemos trabalhando com a variância amostral, dividiríamos pela quantidade de elementos observados subtraída de um (– 1). Nesse exemplo, teríamos:5 dias – 1 = 4 dias.

Vamos então calcular a **variância populacional** para cada funcionário:

Funcionários	Média Aritmética ($\bar{x}$)	
A	$\bar{X}_A = \frac{10 + 9 + 11 + 12 + 8}{5} = \frac{50}{5}$	$\bar{X}_A = 10,0$
B	$\bar{X}_B = \frac{15 + 12 + 16 + 10 + 11}{5} = \frac{64}{5}$	$\bar{X}_B = 12,8$
C	$\bar{X}_C = \frac{11 + 10 + 8 + 11 + 12}{5} = \frac{52}{5}$	$\bar{X}_C = 10,4$
D	$\bar{X}_D = \frac{8 + 12 + 15 + 9 + 11}{5} = \frac{55}{5}$	$\bar{X}_D = 11,0$

Variância → Funcionário A:

$$\text{var}(A) = \frac{(10-10)^2 + (9-10)^2 + (11-10)^2 + (12-10)^2 + (8-10)^2}{5}$$

$$\text{var}(A) = \frac{10}{5} = 2,0$$

Variância →Funcionário B:

$$\text{var}(B) = \frac{(15-12,8)^2 + (12-12,8)^2 + (16-12,8)^2 + (10-12,8)^2 + (11-12,8)^2}{5}$$

$$\text{var}(B) = \frac{26,8}{5} = 5,36$$

Variância →Funcionário C:

$$\text{var}(C) = \frac{(11-10,4)^2 + (10-10,4)^2 + (8-10,4)^2 + (11-10,4)^2 + (12-10,4)^2}{5}$$

$$\text{var}(C) = \frac{9,2}{5} = 1,84$$

Variância →Funcionário D:

$$\text{var}(D) = \frac{(8-11)^2 + (12-11)^2 + (15-11)^2 + (9-11)^2 + (11-11)^2}{5}$$

$$\text{var}(D) = \frac{30}{5} = 6,0$$

Podemos afirmar que a produção diária do funcionário **C** é mais uniforme do que a dos demais funcionários, assim como a quantidade de peças diárias de **D** é a mais desigual. **Quanto maior for a variância, mais distantes da média estarão os valores, e quanto menor for a variância, mais próximos os valores estarão da média.**

Em algumas situações, apenas o cálculo da variância pode não ser suficiente, pois essa é uma medida de dispersão muito influenciada por valores que estão muito distantes da média. Além disso, o fato de a variância ser calculada "ao quadrado" causa uma certa camuflagem dos valores, que dificulta sua interpretação. Uma alternativa para solucionar esse problema é o **desvio padrão**, outra **medida de dispersão.**

O **desvio padrão (dp)** *é simplesmente o resultado positivo da raiz quadrada da variância*. Na prática, o desvio padrão indica qual é o "erro" se quiséssemos substituir um dos valores coletados pelo valor da média. Vamos agora calcular o desvio padrão da produção diária de cada funcionário:

Desvio Padrão →Funcionário A:

$$dp(A) = \sqrt{var\ (A)}$$
$$dp(A) = \sqrt{2,0}$$
$$dp(A) \approx 1,41$$

Desvio Padrão → Funcionário B:team

$$dp(B) = \sqrt{var\ (B)}$$
$$dp(B) = \sqrt{5,36}$$
$$dp(B) \approx 2,32$$

Desvio Padrão →Funcionário C:

$$dp(C) = \sqrt{var\ (C)}$$
$$dp(C) = \sqrt{1,84}$$
$$dp(C) \approx 1,36$$

Desvio Padrão →Funcionário D:

$$dp(D) = \sqrt{var\ (D)}$$
$$dp(D) = \sqrt{6,0}$$
$$dp(D) \approx 2,45$$

Podemos ver a utilização do desvio padrão na apresentação da média aritmética informando o quão “confiável” é esse valor. Isso é feito da seguinte forma:

média aritmética (x) ± desvio padrão (dp)

Se o dono da empresa de nosso exemplo pretende concluir seu relatório com a produção média diária de seus funcionários, ele fará da seguinte forma:

Funcionário A: 10,0 ± 1,41 peças por dia
Funcionário B: 12,8 ± 2,32 peças por dia
Funcionário C: 10,4 ± 1,36 peças por dia
Funcionário D: 11,0 ± 2,45 peças por dia

Exercício

1) Sejam os dados abaixo referentes ao DAP (diâmetro à altura do peito), em cm, de 15 árvores de eucalipto.

10,3 - 10,4 - 10,5 - 11,4 - 11,6 - 11,9 - 12,2 - 12,4 - 12,5 - 13,5 - 13,8 - 14,0 - 14,6 - 15,0 - 15,3

Calcule:

a) desvio padrão
b) coeficiente de variação.

RESP: Desvio Padrão = 1,63 e CV = 12,91%

REFERÊNCIAS

[1] COSTA NETO, PEDRO. L. O. Estatística. Ed. Edgard Blücher, LTDA. 2002.

[2] DEVORE, J. L. Probabilidade e Estatística para Engenharia e Ciências. Ed. Thomson, 2006.

[3] GOODE, WILLIAM J.& HATT, PAUL K. Métodos em Pesquisa Social. Ed. Companhia Editora Nacional, 1979.

[4] HINES, W. & MONTGOMERY, D. C. Probability and Statistics in Engineering and Management Science. Ed. Wiley, 1990.

[5] LAPPONI, J. C. Estatística Usando Excel. São Paulo: Editora Lapponi, 2000.

[6] MENEZES, Alexandre Monteiro, O BIM e o ensino integrado de projeto de edificações. In: COBENGE 2011 - XXXIX Congresso Brasileiro de Educação em Engenharia, 2011, Blumenau.

[7] MONTGOMERY, D. C. & RUNGER, G. C. Estatística Aplicada e Probabilidade para Engenheiros. LTC, 2009

[8] MOOD A. M, GRAYBILL F., BOES, D. C. Introduction to the Theory of Statistics. Editora McGraw-Hill, 1974.

[9] Research gate; Notícias sobre wood frame; disponível em:

[10] <http://www.researchgate.net>Acessoem 20set 2017

[11] SIEGEL, SIDNEY. Estatística Não –Paramétrica. Para Ciências do Comportamento. Ed.McGraw-Hill, 1979.

[12] SIMIELE, D. et al. Substituição Gradativa do agregado natural por agregado reciclado. Tese em Mestrado em Engenharia Civil, São Paulo: UNESP,

[13] WALPOLE, R.; MYERS, R.; MYERS, S & YE, K. Probabilidade e Estatística para Engenharia e Ciências. Ed. Pearson, 2009.

[14] FATURAMENTO mercado pet projeção 2015. ABINPET. 2014. Disponível em: <http://www.agricultura.gov.br/arq_editor/file/camaras_tematicas/Insumos_agropecuarios/79RO/IBGE_PAEB.pdf>. Acesso em: 10 jun. 2016.

[15] www.sindan.org.br